

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО”
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО"
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

ЧАПАЛЮК Богдан Володимирович

УДК 004.8:616-006

ДИСЕРТАЦІЯ

**СИСТЕМИ АВТОМАТИЧНОЇ МЕДИЧНОЇ КОМП'ЮТЕРНОЇ
ДІАГНОСТИКИ З ВИКОРИСТАННЯМ МЕТОДІВ ШТУЧНОГО
ІНТЕЛЕКТУ**

122 — Комп'ютерні науки
Інформаційні технології

Подається на здобуття наукового ступеня
доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Б. В. Чапалюк

Науковий керівник: **Зайченко Юрій Петрович**,
доктор технічних наук, професор

Київ — 2020

АНОТАЦІЯ

Чапалюк Б. В. Системи автоматичної медичної комп'ютерної діагностики з використанням методів штучного інтелекту. — Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 — Комп'ютерні науки. — Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, 2020.

Мета даного дисертаційного дослідження полягає в детальному розгляді, розробці та удосконаленні систем автоматичної комп'ютерної діагностики раку легень використовуючи методи штучного інтелекту, зокрема застосовуючи та удосконалюючи останні досягнення в області глибинного навчання.

Для діагностування раку легень в сучасних медичних закладах використовують комп'ютерну томографію, що представляє собою тривимірне зображення легень пацієнта, отримане за допомогою рентгенівського променю, що пошарово та поступово проходить через тканини людського тіла в різних напрямках, з різних кутів та положень. Такий вид зображень використовується в роботі для аналізу присутності пухлини в легенях за допомогою згорткових нейронних мереж. Однак, такі особливі дані накладають свої складності в розробці систем медичного комп'ютерного діагностування, оскільки при роботі з ними необхідно враховувати їхню тривимірну природу та відповідні просторові зв'язки. Тому, в дисертаційному дослідженні розглядається три основні підходи для роботи з такими даними:

1. Використання двовимірної згорткової нейронної мережі. Для ко-

жного шару КТ знімка застосовується згорткова нейронна мережа. Виходи мережі для кожного шару знімку об'єднуються та фінальний висновок робиться на основі правил навчання за набором зразків.

2. Використання тривимірних згорткових нейронних мереж, які враховують тривимірну природу вхідних даних та можуть віднайти корисні патерни використовуючи всі три просторові вісі. Часто, такі системи розділяють задачу на декілька етапів, кожен з яких використовує тривимірну згорткову нейронну мережу налаштовану під конкретну підзадачу.
3. Використання комбінованої структури двовимірної згорткової та рекурентної нейронних мереж. В такому підході двовимірну згорткову нейронну мережу використовують для представлення вхідних даних в менш мірному просторі шляхом навчання многовиду меншої розмірності. Завдяки цьому на кожному шарі КТ зображення будуть виділятися тільки найбільш важливі високорівневі ознаки. Отримані ознаки обробляються двонаправленою рекурентною нейронною мережею з вентиляним вузлом (англ. *bidirectional gated recurrent neural network*), яка навчається складним нелінійним функціям, що описують просторові залежності та вплив між ними. Вихід рекурентної мережі повертає ймовірність наявності пухлини на знімку.

В рамках даного дисертаційного дослідження проводиться аналіз та виконується експерименти для кожного підходу, а отримані результати порівнюються з роботами інших авторів. Експерименти показують, що найбільш точними є системи побудовані із декількох тривимірних згорткових нейронних мереж (одна мережа сегментує потенційні проблемні регіони, інша класифікує присутність в таких регіонах пухлини). Однак, такі системи мають дуже великі обчислювальні вимоги, через те що використовують

операцію тривимірної згортки, вимоги до обчислювальної потужності якої ростуть кубічно зі збільшенням розмірності вхідного зображення. В такому випадку, запропонована архітектура рекурентної згорткової нейронної мережі дозволяє отримати точність роботи системи на достатньо високому рівні, в той же час використовуючи значно менш вимогливу до обчислювальних потужностей та пам'яті операцію двовимірної згортки.

Наукова новизна отриманих результатів дисертації полягає в запропонованому здобувачем методі побудови комбінованої структури системи комп'ютерної діагностики, що полягає в поєднанні двовимірної згорткової та двонаправленої рекурентної нейронної мережі LSTM. На відміну від інших рішень, така система враховує просторові зв'язки між різними шарами знімку комп'ютерної томографії шляхом використання двонаправленої рекурентної нейронної мережі, на вході якої використовують високорівневі ознаки сформовані за допомогою двовимірної згорткової нейронної мережі. Високорівневі ознаки будуються для кожного шару знімку пацієнта. За результатами експериментів така архітектура нейронної мережі змогла досягти значення AUC ROC на рівні 83%, що трохи нижче у порівнянні з системами тривимірних згорткових нейронних мереж, що показують значення AUC ROC на рівні 90-95%. Однак, отримані результати є найвищими результатами для рекурентних нейронних мереж, що застосовуються для побудови систем комп'ютерної діагностики раку легенів. Також, запропонована архітектура має вищу швидкодію, що досягається шляхом використання операції двовимірної згортки замість операції тривимірної згортки, вимоги якої до обчислювальної потужності та пам'яті ростуть квадратично з розміром вхідних даних, а не кубічно.

Для ефективного навчання комбінованої структури згорткової рекурентної нейронної мережі був запропонований механізм м'якої уваги, що надав можливість нейронній мережі отримати інформацію про локацію пухлини під час навчання. Згідно проведених експериментів, такий підхід допоміг

покращити показники метрики AUC ROC більш ніж на 8%.

Практичне значення отриманих результатів полягає в розширенні та удосконаленні існуючих методів побудови систем комп'ютерної діагностики. Запропонована комбінована структура згорткової нейронної мережі та двонаправленої рекурентної мережі дозволяє отримати достатньо високу точність роботи системи та підвищує точність роботи системи у порівнянні з використанням звичайних рекурентних нейронних мереж. Також, така система відзначається використанням меншої кількості ресурсів чим у тривимірної згорткової нейронної мережі. Проведені експерименти та аналіз існуючих методів систем комп'ютерної діагностики дозволив сформулювати необхідні вимоги та підходи, які потрібно використовувати в залежності від пріоритету швидкодії чи точності роботи системи. Запропонований механізм м'якої уваги дозволяє значно підвищити ефективність навчання комбінованих архітектур згорткових рекурентних нейронних мереж.

Результати дисертаційного дослідження впроваджено в НДР за темою "Розроблення та дослідження методів обробки, розпізнавання, захисту та зберігання медичних зображень в розподілених комп'ютерних системах" за номером держ реєстрації 0117U004267 (тема №2021п, код КВНТД І.1 01.05.02). Також, основні результати роботи викладені в 6 друкованих наукових роботах, з них двоє статей в наукових фахових виданнях України, 2 опубліковано в іноземних журналах, що індексується в Google Scholar та інших базах даних, 1-а стаття у виданні, що входить до Web of Science Core Collection та SCOPUS. Також опубліковано одну роботу в тезах доповідей міжнародної наукової конференції.

Ключові слова: глибинне навчання, згорткова нейронна мережа, рекурентна нейронна мережа, механізм уваги, тривимірна згортка, система автоматичної медичної комп'ютерної діагностики раку легенів.

ABSTRACT

Chapaliuk B. V. Computer-aided diagnostic systems with artificial intelligence methods usage. — Qualification scientific work in the form of manuscript.

Thesis for doctor of philosophy degree in speciality 122 — Computer science. — National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, 2020.

The primary purpose of the thesis research is to consider and analyse, develop, and improve computer-aided detection systems for lung cancer using artificial intelligence methods. In particular, work examines, applies, and improves the latest achievements in the field of deep learning.

The modern medicine practice recommends using low-dose computed tomography (also called CT scan) of the human chest to investigate lung cancer presence. CT scan represents a three-dimensional image of the patient’s lungs obtained by the X-ray that gradually passes through the human body’s tissue, layer by layer, in different directions, angles, and positions. Such images are usually used for computer-aided lung cancer detection systems based on convolution neural networks. However, such special data impose additional difficulties in building and implementing computer-aided diagnosis systems because implementation should consider the three-dimensional nature of the data and related spatial dependencies. That’s why thesis research considers three main approaches for working with such data types:

1. The first approach uses a two-dimensional convolution neural network.

In this approach, the convolution neural network is applied for each layer in the CT scan, and network outputs are combined by predefined

rules. The final inference is based on the rules of multiple-instance learning.

2. The second approach uses three-dimensional convolution neural networks that take into account the three-dimensional data nature, which can learn the pattern among all three dimensions. Often, systems that use 3D convolutional neural networks divide the problem into several stages. In this case, each step was responsible and configured for a specific subtask.
3. The third approach uses a combined structure of convolution and recurrent neural networks. In this case, the convolution neural network is used to represent input data in the lower dimensional space by learning the manifold in the lower dimensional space. Due to this, each processed CT scan layer represents only the most useful high-level features. These high-level features are examined and combined by the bidirectional gated recurrent neural network that learned complex patterns. The patterns describe dependencies and spatial relationships that can be used to diagnose lung cancer. The output of recurrent neural networks returns the probability of lung cancer presence on the CT scan.

In the scope of the dissertation research, each approach is investigated and analysed in detail. For each approach, experiments were created for the approach checking, compared to the results of works with other authors. The experiment results show the most precise inference has the systems based on the several three-dimensional neural networks (one network segments potentially dangerous pulmonary nodules; meanwhile, the second network classifies the presence of cancer in the segmented regions). However, such neural networks have huge computational power and memory requirements. The reason is the usage of three-dimensional convolution operation, in which computation power requirements grow cubically when the input image dimension

increases. Conversely, the proposed structure of the convolution and recurrent neural network show results that are precise enough and meanwhile, use a much less computationally and memory demanding operation: the two-dimensional convolution operation.

The scientific novelty of the thesis results is the creation of the methods to build a combined structure of the two-dimensional convolutional neural network and bidirectional recurrent neural network LSTM. Unlike other solutions, such a system takes into account spatial relationships between different CT scan layers by using bidirectional recurrent neural networks. The input of the recurrent neural network consists of the high-level features retrieved from the two-dimensional convolution neural network. The high-level features are built for each layer. The experiment results show that such a neural network achieves 83% of AUC ROC metric. That number is lower than the number three-dimensional neural networks show, which can reach 90–95% of AUC ROC metric. Nevertheless, obtained results are the highest results that were achieved for recurrent neural networks used for lung cancer detection systems. Also, the proposed neural network architecture has higher performance characteristics, which is achieved by using a two-dimensional convolution operation instead of a three-dimensional convolution.

In addition, the soft-attention mechanism was proposed to improve the learning process effectiveness of the combined structure of convolutional and recurrent neural networks. The proposed mechanism of explicit attention gives the neural network information about the cancer location during the training process. Per the experiments, the proposed mechanism improved the AUC ROC metrics value by more than 8%.

The practical significance of the obtained results is the ability to extend and improve existing methods of building computer-aided lung cancer diagnosis systems. The proposed combined structure of the convolutional and bidirectional recurrent neural networks allows for achieving a high enough

precision level of system work and increasing system precision compared to the simple recurrent neural network usage. Also, the proposed architecture uses much less memory and has lower computation requirements than three-dimensional neural networks. Conducted experiments and analysis of existing methods for building computer-aided lung cancer detection systems enable formulating requirements and approaches that should be used depending on the priority of the performance or precision. The proposed soft attention mechanism makes the learning of the convolutional recurrent neural network more effective.

The results of the dissertation research were captured in the scientific research work “Development and Study of Methods of Medical Images Processing, Recognition, Protection and Storing in Distributed Computer Systems”(Ref. No. 2021p, Reg. No. 0117U004267). Also, the main results were published in six scientific work print pieces; two of the papers were articles in Ukrainian scientific professional publications, two papers were in a foreign scientific journals that are indexed in Google Scholar and other scientific databases, and another paper was in a magazine indexed in the Web Of Science Core Collection and Scopus. In addition, there is an article published in international conference thesis collection.

Key words: deep learning, convolution neural network, recurrent neural network, attention, three-dimensional convolution, computer-aided lung cancer detection system.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. B. V. Chapaliuk and Y. P. Zaychenko, “Автоматична медична діагностика на базі зображень комп’ютерної томографії,” in *System Analysis and Information Technology (SAIT)*. Kyiv: ESC IASA NTUU Igor Sikorsky Kyiv Polytechnic Institute, 2018, p. 155.
2. Bohdan V. Chapaliuk and Yuriy P. Zaychenko, “Огляд методів сегментації медичних зображень,” *Системні дослідження та інформаційні технології*, no. 1, pp. 72–81, 2018.
3. B. Chapaliuk and Y. Zaychenko, “Deep learning approach in computer-aided detection system for lung cancer,” in *2018 IEEE First International Conference on System Analysis Intelligent Computing (SAIC)*, Oct 2018, pp. 1–4.
4. B. Chapaliuk and Y. Zaychenko, “End-to-end deep learning strategies for computer-aided lung cancer detection systems,” *SCIREA Journal of Mathematics*, vol. 4, no. 5, pp. 140–155, 2019.
5. B. V. Chapaliuk and Y. P. Zaychenko, “Використання рекурентних нейронних мереж для автоматичної діагностики раку легенів,” *Системні дослідження та інформаційні технології*, no. 3, pp. 33–40, 2019.
6. B. Chapaliuk, “Overview of the three-dimensional convolutional neural networks usage in medical computer-aided diagnosis systems,” *American Journal of Neural Networks and Applications*, vol. 6, no. 2, pp. 22–28, 2020.

ЗМІСТ

Вступ	13
Розділ 1. Огляд існуючих рішень. Структура та методи систем комп'ютерної діагностики	20
1.1. Системи комп'ютерної діагностики	22
1.2. Багато-атласні методи сегментації медичних зображень . . .	24
1.2.1. Не параметричні багато-атласні методи	26
1.2.2. Параметричні або ймовірнісні багато-атласні методи .	36
1.3. Методи глибинного навчання для сегментації та класифікації	36
1.4. Підсумки та постановка задачі дослідження	46
Розділ 2. Тривимірні нейронні мережі для автоматичної діагностики раку легенів	50
2.1. Згорткова нейронна мережа та її тривимірна версія	51
2.1.1. Операція згортки. Використання операції згортки для обробки зображень	52
2.1.2. Властивості тривимірної згорткової нейронної мережі	56
2.1.3. Операція субдискретизації. Її властивості та використання	60
2.1.4. Навчання та регуляризація тривимірних згорткових нейронних мереж	62
2.2. Використання тривимірної згорткової нейронної мережі для діагностування раку легенів	70
2.3. Висновки до розділу	78
Розділ 3. Рекурентні нейронні мережі для діагностування раку легенів	79

3.1. Рекурентні нейронні мережі	80
3.2. Двонаправлені рекурентні нейронні мережі	86
3.3. Рекурентні нейронні мережі з вентиляним вузлом	88
3.4. Механізм уваги в задачах комп'ютерного зору	92
3.5. Застосування рекурентної згорткової нейронної мережі з механізмом уваги для діагностики раку легенів	97
3.6. Висновки до розділу	101
Розділ 4. Експериментальні дослідження архітектур нейронних мереж для обробки КТ зображень легенів	102
4.1. Діагностика раку легенів. Опис даних	102
4.2. Попередня обробка КТ зображень	107
4.3. Збільшення даних	110
4.4. Двовимірна згортка та навчання за набором зразків	111
4.5. Рекурентна нейронна мережа з механізмом уваги	115
4.6. Одноетапні тривимірні згорткові нейронні мережі	116
4.7. Підсумки отриманих результатів	118
Висновки	122
Список використаних джерел	126

ВСТУП

Обґрунтування вибору теми дослідження. Медична діагностика та медицина в цілому є одними із найбільш складних та важливих галузей людства. Накопичення медичних знань про організм людини, патології, симптоми та ознаки різних захворювань, можливість швидкої та автоматичної обробки такої інформації є надважливими завданнями, вирішуючи які, можливо врятувати велику кількість життів. Це є причиною того, чому багато науковців та інженерів вкладають багато зусиль, намагаючись покращити та пришвидшити медичне обслуговування та діагностування шляхом розробки різноманітних автоматизованих комп'ютерних систем. Активний розвиток таких систем дозволяє пришвидшити та спростити діагностику серйозних захворювань. Завдяки цьому у лікарів з'явиться можливість діагностувати захворювання на більш ранній стадії, що в складних випадках може грати суттєву роль в процесі лікування пацієнта. Як відомо, навіть однакові хвороби можуть протікати по-різному у різних людей і можуть мати різні наслідки. Тому, персоналізована медицина може значно покращити медичне обслуговування для кожної людини, а також зменшити кількість помилок в лікуванні та діагностуванні.

В рамках даної роботи розглядається проблема побудови системи автоматичної діагностики раку легенів. Рак легенів є однією із найбільш розповсюджених причин смерті пацієнтів, що мають проблеми з онкологією. Причиною цьому є агресивність та складність протікання хвороби, виявлення патологій на пізніх стадіях. Відповідно до даних Всесвітньої Організації Охорони Здоров'я (англ. World Health Organization), протягом 2012 року було зафіксовано близько 1.8 мільйонів нових випадків раку легенів та близько 1.6 мільйонів випадків смертей пов'язаних з ним. Саме через

велику кількість летальних випадків у пацієнтів рак легенів вважається одним із найбільш агресивних видів ракових захворювань. Протягом п'яти років після діагностування раку легенів, виживає тільки 10-15% пацієнтів. Але, якщо рак вдається діагностувати на ранніх стадіях, то протягом десяти років виживає близько 75% відсотків пацієнтів [1].

З плином часу сумна статистика захворюваності раку легенів тільки погіршується. Згідно з оновленими даними від Всесвітньої Організації Охорони Здоров'я [2], за 2018 рік по всьому світі було зареєстровано 2.09 мільйонів випадків захворюваності на рак легенів, з який 1.76 мільйона випадків були смертельними. Для порівняння, кількість зареєстрованих випадків раку молочної залози також складало близько 2.09 мільйонів. Однак, з них смертельними виявилися тільки 627 000 випадків, що значно менше у порівнянні з раком легенів. В Україні статистика захворюваності раку легенів теж не втішна. Так, згідно даних міністерства охорони здоров'я України [3] на території держави фіксується близько 13 тисяч нових випадків раку легенів на рік, більшість яких пов'язаних з шкідливими звичками, як-от паління.

Існує два основних види раку легенів [4]: рак легенів центральної форми, що розвивається в центральних бронхах та рак легенів периферичної форми, що локалізується в бронхіолах та в паренхімі легень (легенева тканина, альвеоли). Рак легенів центральної форми зазвичай діагностується на ранніх стадіях, через те що симптоматика проявляється ще на ранній стадії. Більш небезпечним вважається рак легенів периферичної форми, оскільки така форма захворювання може довгий час протікати без специфічних симптомів. Тому, основним способом боротьби з раком легенів є періодична діагностика шляхом виконання флюорографії або комп'ютерної томографії. Важливо, щоб такі перевірки проводилися регулярно та аналізувалися разом з минулими знімками. В такому випадку розробка автоматичних систем діагностики раку легенів може допомогти пришвидшити,

полегшити, стандартизувати та здешевити процес знаходження патологій. Така система має вміти не тільки діагностувати проблему, а й локалізувати пухлину на знімку для пояснення причин висновків такої системи. В такому випадку система зможе допомогти лікарям робити більш якісну та швидку діагностику.

В останні роки технології штучного інтелекту здійснили серйозний прорив в таких областях як комп'ютерне бачення, обробка текстової інформації, розпізнавання мови та інших. Це стало можливим завдяки використанню глибоких нейронних мереж та методів глибокого навчання. Подібні ж методи можливо використати для автоматизації та покращення роботи систем медичної діагностики, в тому числі у області діагностування раку легенів.

Таким чином, існує велика потреба у автоматизації, стандартизації, пришвидшенні та підвищенні якості діагностики раку легенів шляхом створення автоматизованих систем асистування лікарям. Це зумовлює актуальність розробки нових методів штучного інтелекту для систем автоматичної діагностики раку легенів.

Мета і завдання дослідження. Метою дослідження є детальний розгляд, розробка та удосконалення систем автоматичної комп'ютерної діагностики раку легенів на основі методів штучного інтелекту, зокрема використовуючи останні досягнення в області глибокого навчання.

Для досягнення поставленої мети, були поставлені наступні завдання дослідження:

- Дослідити та виконати аналіз існуючих методів побудови систем автоматичного комп'ютерного діагностування.
- Дослідити існуючу структуру систем автоматичної діагностики раку легенів.
- Дослідити архітектури та види глибоких нейронних мереж які ви-

користовуються для роботи з медичними зображеннями.

- Дослідити, розробити або модифікувати існуючі архітектури згорткових нейронних мереж для задач сегментації та класифікації.
- Розробити власну систему автоматичного комп'ютерного діагностування раку легенів.
- Виконати експериментальні дослідження розроблених моделей і методів для визначення їх ефективності.

Об'єкт дослідження - процеси діагностики раку легенів.

Предмет дослідження - моделі та методи побудови структур згорткових та рекурентних нейронних мереж для автоматичної системи комп'ютерної діагностики раку легенів.

Методи дослідження. Методи статистичної обробки експериментальних даних, теорія штучних нейронних мереж та глибинного навчання, методи сегментації, методи обробки зображень. Дані методи були вибрані для виконання поставлених цілей з огляду на те, що вони є найбільш популярними та добре себе зарекомендували в системах комп'ютерної медичної діагностики. Також, методи глибинного навчання показують найбільш багатобічючі результати в задачах пов'язаних з обробкою зображень.

В процесі дослідження використовувалися фреймворки машинного навчання keras та tensorflow, мова програмування python.

Наукова новизна отриманих результатів. Наукова новизна отриманих результатів полягає в:

- Вперше розроблена нова структура автоматичної системи діагностики раку легень на основі згорткових нейронних мереж, яка на відміну від існуючих систем дає можливість обробляти тривимірні дані, враховувати тривимірну структуру знімків рентгенівської комп'ютерної томографії легенів пацієнта і забезпечує значне підвищення точності діагностики.

- Запропоновано метод побудови комбінованої структури системи автоматичної діагностики, що полягає в інтеграції двовимірної згорткової та рекурентної нейронних мереж LSTM. На відміну від відомих рішень, запропонована система враховує просторові зв'язки між різними шарами знімка комп'ютерної томографії шляхом використання рекурентної нейронної мережі, на вході якої використовуються ознаки високого рівня, сформовані за допомогою двовимірної згорткової нейронної мережі; при цьому ознаки високого рівня будуються для кожного шару знімка пацієнта. Запропонований підхід дав можливість суттєво підвищити швидкодію діагностичної системи.
- Запропоновано механізм м'якої уваги для комбінованої структури згорткової та рекурентної нейронних мереж, який на відміну від відомих підходів дає можливість використати анотовану інформацію про локацію злоякісних утворень наявних у вибірках даних. З використанням запропонованого механізму уваги, нейронна мережа під час навчання напрямую отримує інформацію про локацію утворень, що мають бути виділені на конкретному шарі знімка пацієнта. Такий підхід дозволив значно підвищити точність роботи діагностичної системи в цілому.

Практичне значення отриманих результатів. Отримані результати експериментальних досліджень розширюють та удосконалюють методи аналізу даних для систем комп'ютерної діагностики. Запропонована комбінована структура згорткової нейронної мережі та двонапрямної рекурентної мережі дають можливість отримати високу точність функціонування системи та підвищують точність роботи системи порівняно з використанням звичайних рекурентних нейронних мереж. Також, запропонована система має перевагу стосовно використання меншого об'єму обчислюваль-

них ресурсів ніж тривимірні згорткові нейронні мережі. Проведені експерименти та аналіз існуючих методів систем комп'ютерної діагностики дозволив сформулювати необхідні вимоги та підходи, які потрібно використовувати залежно від пріоритету – швидкодії чи точності функціонування системи. Запропонований механізм м'якої уваги дозволяє значно підвищити ефективність навчання комбінованих архітектур згорткових рекурентних нейронних мереж, що дає можливість значно підвищити точність діагностики.

Теоретичні і практичні результати роботи корисні для застосування у навчальному процесі технічних університетів, зокрема, вони можуть бути використані в курсах «розпізнавання образів», «теорія прийняття рішень».

Особистий внесок здобувача. Усі основні результати дисертаційного дослідження, представлені до захисту, одержані автором особисто. У публікаціях у співавторстві, здобувачеві належить: виконання адаптації двовимірних згорткових нейронних мереж до тривимірних вхідних даних шляхом застосування тривимірної операції згортки та побудова система автоматичної діагностики раку легенів на основі адаптованих нейронних мереж [5]; розробка комбінованої структури рекурентної та двовимірної згорткової нейронних мереж з покращеним механізмом уваги [6]; аналіз та порівняння результатів отриманих здобувачем та іншими авторами при розробці систем автоматичної діагностики раку легенів [7]. В роботі [8] здобувачем була проведено дослідження та аналіз систем автоматичної медичної діагностики, що використовувалися перед набуттям популярності методів глибинного навчання та порівняння їхніх переваг і недоліків із сучасними підходами.

Апробація матеріалів дисертації. Основні положення та отримані наукові результати, що викладені в даній дисертаційній роботі, пройшли апробацію на різних науково-технічних конференціях: міжна-

родній науково-технічній конференції “System Analysis and Information Technology 2018” (SAIT 2018), IEEE конференції “Системний аналіз та інтелектуальні обчислення 2018” (SAIC 2018).

Результати роботи впроваджувалися в НДР за темою “Розроблення та дослідження методів обробки, розпізнавання, захисту та зберігання медичних зображень в розподілених комп’ютерних системах” за номером держреєстрації 0117U004267 (тема №2021п, код КВНТД І.1 01.05.02) та виконувалась в 2017-2019 роках в НТУУ КПІ ім. Сікорського на кафедрі прикладної математики.

Публікації. Основні результати роботи викладені в 6 друкованих наукових роботах, з них двоє статей в наукових фахових виданнях України, 2 опубліковано в іноземних журналах, що індексується в Googel Scholar та інших базах даних, 1-а стаття у виданні, що входить до Web of Science Core Collection та SCOPUS. Також опубліковано одну роботу в тезах доповідей міжнародної наукової конференції.

Структура та обсяг дисертації. Дисертації складається зі вступу, 4 розділів, висновків та списку використаних джерел (123 найменування). Загальний обсяг дисертаційної роботи складає 139 сторінок.

РОЗДІЛ 1

ОГЛЯД ІСНУЮЧИХ РІШЕНЬ. СТРУКТУРА ТА МЕТОДИ СИСТЕМ КОМП'ЮТЕРНОЇ ДІАГНОСТИКИ

Розвиток та застосування цифрових технологій в сучасному світі має масштабний характер, що призводить до тенденції перенесення всієї інформації на цифрові носії. Така сама тенденція відчувається і в медичній сфері. В розвинутих країнах лікарні будують свої власні бази даних медичної інформації пацієнтів, відомі як системи електронних медичних записів (англ. electronic health records). Наприклад, США почала розвивати систему електронних медичних записів ще з 2009 року і на даний момент всі основні компанії, що надають медичні послуги, використовують подібні цифрові системи [9]. В Україні систему електронних медичних записів тільки починають впроваджувати на державному рівні. Але, сучасна цифрова діагностика давно доступна у приватних клініках.

Доступність даних у цифровому вигляді дозволяють будувати сучасні складні медичні системи, що обробляють доступну інформацію та допомагають, або спрощують діагностику захворювань. В даній роботі розглядається можливості цифрової діагностики, що виконується на основі зображень комп'ютерної томографії та використовує сучасні методи штучного інтелекту, зокрема підходи глибинного навчання.

Аналіз цифрових зображень внутрішніх органів людини є надзвичайно складною задачею. Основною причиною цього є те, що анатомічні об'єкти організму мають велику варіацію в формі та зовнішньому вигляді через велику кількість факторів:

- *Шуми та артефакти сенсора.* Як і будь-який фізичний сенсор, медичне обладнання генерує шум та артефакти через природу

сенсора та спосіб формування знімку. Наприклад, високодозова комп'ютерна томографія (КТ) отримує зображення з значно меншою кількістю артефактів, ніж низькодозова КТ. Однак, низькодозові КТ знімки мають пріоритет при проведенні досліджень через їх менший негативний вплив на організм людини.

- *Різна форма тіла пацієнтів.* Кожна людина має власну форму тіла. Людина може бути повною, худю, високою та інше. Всі ці фактори впливають на форму внутрішніх органів, та вносять певну варіативність при проведенні автоматичної діагностики.
- *Доступність зображення тільки частини тіла людини.* Доза радіації є суттєвим аргументом при проведенні діагностики людини шляхом рентгенівського випромінювання. Щоб зменшити дозу радіації, яку отримує людина при обстеженнях, зазвичай призначаються знімки локальної частини тіла. В таких випадках ми втрачаємо глобальну інформацію про внутрішній стан організму людини, що може бути важливим при виявленні патологій та діагностуванні.
- *Схожі властивості тканини органів людини.* Всі людські органи складаються з певного м'якого матеріалу. Однак, в деяких випадках властивості тканини різних органів можуть бути подібними, що ускладнює їхній аналіз на зображенні КТ.

У цьому розділі буде розглянута інформація про структуру та типи комп'ютерних систем медичної діагностики, а також методи, що використовуються для обробки та аналізу складних медичних даних. Будуть розглянуті багато-атласні методи сегментації медичних зображень, що розбивають задачу сегментації на декілька етапів для спрощення роботи системи. Також, ми розглянемо сучасні методи глибинного навчання, що можуть використовуватися як частина багато-атласних методів, або замінити повністю всі етапи однією моделлю. В кінці розділу буде описано переваги і недоліки існуючих систем та постановка задачі дисертаційної роботи.

1.1. Системи комп'ютерної діагностики

Під системами комп'ютерної діагностики розуміють такі системи, що допомагають та асистують лікарям при інтерпретації медичних зображень для діагностування захворювань у пацієнтів. Зазвичай, як медичні зображення можуть використовуватися рентгенівські знімки, рентгенівська комп'ютерна томографія, магнітно-резонансна томографія та зображення отримані ультразвуковою діагностикою, що надають інформацію про внутрішню структуру органів людини та використовуються для визначення патологій в організмі пацієнта. Системи комп'ютерної діагностики поділяються на два основні типи [10]:

1. Системи комп'ютерної діагностики, що детектують та помічають області на медичному зображенні. Такі області відображають потенційну аномалію, що може бути причиною захворювання. Основною задачею такої системи є зменшення навантаження на рентгенолога шляхом автоматичної детекції необхідних областей, а також додаванням опису знайденої аномалії.
2. Системи комп'ютерної діагностики, для яких основною метою є діагностування захворювання використовуючи наявні медичні зображення пацієнта. Ціллю таких систем є повністю автоматична діагностика захворювань, що в ідеальному випадку немає вимагати втручання лікаря. На практиці результати діагностування таких систем все ще вимагають валідацію зі сторони експерта, тому часто такі системи надають додаткову інформацію про причини прийнятого рішення, наприклад інформацію про локацію виявленої аномалії.

Комп'ютерні системи діагностики є важливою частиною сучасної медицини та використовуються для зменшення навантаження на лікарів, прискорення та покращення якості діагностування захворювань у пацієнтів. В деяких випадках, швидкість діагностування патологій в організмі люди-

ни може врятувати життя пацієнта, оскільки надасть можливість лікарям вчасно призначити та виконати необхідні процедури. Однак, такі системи є дуже складними для побудови та реалізації через непросту структуру та внутрішню будову організму людини. Проблема розробки такої системи ускладнюється ще більше через відсутність великої кількості даних, що особливо важливо для систем на базі машинного навчання. Тому, часто пайплайн системи комп'ютерної діагностики складається із декількох основних етапів [11, 12]:

1. Етап попередньої обробки медичного зображення. Даний етап залежить від типу даних та обладнання, що використовується для отримання медичного зображення. Основною метою даного етапу є видалити дефекти, що пов'язані з процесом отримання таких зображень. Прикладом кроків попередньої обробки може бути видалення шумів, покращення контрастності зображення, відфільтровування непотрібних елементів та конвертування зображення в зручний для наступних алгоритмів формат.
2. Етап сегментації зображення. Мета даного етапу полягає у виділенні регіону або органу який необхідно аналізувати з поміж інших органів та тканин для зменшення вимог до обчислювальної потужності на наступних етапах. Метод сегментації залежить від типу зображення та медичного домену. Наприклад, для системи діагностики раку легенів на даному етапі виконуватиметься процедура виділення легенів з поміж інших тканин, таких як кістки, судини та інші.
3. Етап виділення областей інтересу (англ. region of interest). Мета цього етапу полягає в детектуванні аномальних областей сегментованого органу. Часто знайдені регіони аналізуються на набір спеціальних морфологічних характеристик, що можуть бути корисним для діагностування того чи іншого захворювання. Часто, така інформація

надається кінцевому користувачеві такої системи.

4. Етап класифікації. Даний етап виконує індивідуальний аналіз виділених областей інтересу, та повертає ймовірність присутності ознак захворювання на зображенні.

В більш ранніх підходах створення систем комп'ютерної діагностики етап класифікації не виконувався, оскільки загальна точність виділення аномалій на медичному зображенні була не високою, а визначення фінального діагнозу покладалося на експерта. Експерт проглядав всі виділені системою області, а також шукав можливі області, що такий алгоритм міг пропустити. В більш сучасних системах, що базуються на методах глибокого навчання, точність роботи системи є значно вищою, тому науковці та інженери почали реалізовувати більш просунуті системи, які виконують фінальний етап класифікації та роблять фінальний висновок про присутність захворювання.

В наступних підрозділах розглянемо популярні протягом останніх 15-ти років методи сегментації та виділення областей інтересу, що використовувалися в системах комп'ютерної діагностики. Так, в 2000-х роках найбільш популярними методами сегментації зображення були багато-атласні методи, що могли використовуватися як для другого, так і для третього етапів. Також, будуть розглянути методи сегментації та класифікації, що є найбільш популярними в останні роки та використовують методи машинного навчання.

1.2. Багато-атласні методи сегментації медичних зображень

Під багато-атласними методами зазвичай розуміють підходи, в межах яких автоматична сегментація анатомічної структури на зображенні відбувається шляхом розповсюдження набору міток з відомого набору зображень атласів на координати схожих структур нового зображення. Процес

розповсюдження міток з набору зображень на структури в новому зображенні називають процесом реєстрації зображення. Зазвичай, зображення атласи створюються за допомогою спеціалізованого інтерактивного програмного забезпечення, використовуючи яке, експерти створюють анотації в ручному режимі [13, 14]. Зазвичай такими експертами виступають медичні працівники або рентгенологи, робота яких є дорогою.

Багато-атласні методи поділяють на два основні типи - ймовірнісні або параметричні методи та непараметричні методи. Під параметричними методами розуміють алгоритми, які моделюють ймовірнісне представлення анотованих зображень атласів, або іншими словами будують ймовірнісну модель розподілу зображень атласів [15, 16]. Під непараметричними методами зазвичай розуміють алгоритми, що використовують наявний набір зображень атласів напряду використовуючи неймовірнісні алгоритми [14, 17]. Ті методи, які використовують декілька атласів і називають багато-атласними методами, через те, що при своїй роботі вони використовують вибірку з багатьох еталонних зображень.

Перед початком активного розвитку методів глибинного навчання, що використовують глибинні нейронні мережі для моделювання ймовірнісного розподілу зображень атласів, найбільш популярними та точними вважалися не параметричні багато-атласні методи [18]. В той час існуючі ймовірнісні моделі будували достатньо слабку ймовірнісну модель набору зображень атласів, тоді як не параметричні методи дозволяли використовувати та вибирати найбільш відповідну до цільового зображення підмножину зображень атласів. На той момент вважалося, що за рахунок використання цілого набору найбільш відповідних зображень можна досягти значно кращої роботи моделі за рахунок більшої варіативності. Але серйозним недоліком такого підходу є значні вимоги до обчислювальної потужності системи. Тож, для побудови систем автоматичної сегментації медичних зображень використовували ймовірнісні методи у випадку коли була більш важлива

швидкодія, а непараметричні методи коли пріоритетом було точність та коректність роботи. Розглянемо багато-атласні непараметричні методи більш детально.

1.2.1. Не параметричні багато-атласні методи. Формально, метою не параметричного багато-атласного методу є сегментація цільового медичного зображення T використовуючи набір зображень атласів A_1, \dots, A_m та їхні відповідні мітки L_1, \dots, L_m . Значення міток може приймати значення 1 або 0 в залежності від того чи воксель (чи піксель у випадку двовимірного зображення) x належить до заданої структури чи ні. Щоб досягнути цього, багато-атласні методи включають в себе наступні три основні етапи:

1. Вибір зображення атласу.
2. Реєстрація зображення.
3. Злиття мітки.

Також, в більш сучасних алгоритмах можуть використовуватися певні додаткові етапи. Наприклад, пайплайн багато-атласного методу можуть складатися з етапу онлайн навчання [19, 20, 21, 22] та кроків пост обробки результатів [23, 24, 25, 26].

1.2.1.1. Етап вибору зображення атласу. В рамках етапу вибору зображень атласів (англ. atlas selection), багато-атласний алгоритм підбирає підмножину зображень атласів, що найбільш анатомічно схожі до цільового об'єкта S_T в наборі зображень атласів $1 \dots m$. Тобто, ми вибираємо об'єкт, який ми намагаємося знайти в цільовому зображенні, та вибираємо певну кількість зображень атласів із наявного набору даних, який містять цей об'єкт в схожій формі як на цільовому зображенні. В рамках цієї процедури не використовуються всі наявні зображення атласи по декільком причинам:

- Багато систем автоматичної діагностики мають специфічні вимоги до швидкодії та обчислюваних ресурсів, що обумовлює шукати шляхи до більш ефективної роботи системи. Для деяких медичних систем швидкодія грає надважливу роль, через те що в деяких випадках результати діагностики мають приходити в найкоротші терміни для оперативного втручання і запобігання смерті пацієнта.
- Виключаючи зображення атласи, які не є релевантними для сегментації необхідного об'єкта на цільовому зображенні, підвищується загальна коректність та точність роботи системи. Справа в тому, що додаткові нерелевантні зображення будуть мати негативний вплив на знаходження об'єкту, оскільки будуть зміщувати та спотворювати об'єкт, який необхідно виявити на зображенні.

Коли тільки почали свій розвиток непараметричні багато-атласні методи, підмножина зображень атласів вибиралася випадковим чином [17]. При такому виборі зображень атласів, коректність та точність роботи всієї системи зростала при збільшенні кількості зображень в підмножині. Але покращення відбувалося до певної межі та при додаванні занадто великої кількості зображень, коректність та точність роботи системи починала падати через присутність невідповідних кандидатів. При використанні декількох випадкових зображень система працювала краще через те, що поєднання декількох репрезентативних зображень допомагало нівелювати помилки або неточності в одному індивідуальному зображенні атласі. Ця ідея є фундаментальною для багато-атласних методів.

Випадковий підбір зображень атласів має суттєвий недолік - є достатньо висока ймовірність, що випадково вибрані зображення атласи, або деякі з них, не будуть репрезентативними для об'єкта цільового зображення атласу. Щоб такої проблеми не було, потрібно відфільтровувати всі зображення атласи, що містять не відповідні анатомічні характеристики. Тому, основна задача яку вирішує етап вибору зображень атласів полягає

в визначені та розробці функції, що найкраще відображає подібність між вибраним атласом A_i та цільовим зображенням T . Таку функцію в алгоритмах багато-атласних методів називають метрикою подібності між двома атласами. В залежності від поставленої задачі та даними, для яких розроблюється алгоритм, метрикою подібності можуть бути вибрана інтенсивність значень вокселів зображення (або пікселів, якщо працюють з двовимірними даними) або функція нормалізованої взаємної інформації (англ. *normalizaed mutual information*) [27], функція консистентності реєстрації (англ. *registration consistency*)[28], взаємокореляційна функція значень інтенсивності (англ. *cross-correlation of intensity*) вокселів, або пікселів, якщо система працює з двовимірними даними [29].

В більш нових роботах були запропоновані більш просунуті техніки вибору зображень атласів. Наприклад, робота [24] змогла підвищити коректність та точність роботи системи використавши алгоритм кластеризації. В рамках роботи автори запропонували визначити вектор узгодженості між двома атласами та застосувати алгоритм кластеризації методом k -середніх до заданого вектору. Крім того, застосовувався параметр ранжування, що оцінює суму різниці квадратів між середніми значеннями отриманих кластерів. Цей параметр ранжування використовувався до результуючих кластерів, отриманих алгоритмом кластеризації. Такий підхід допоміг підвищити коректність та точність роботи системи у порівнянні із підходами, описаними вище.

З іншого боку, етап вибору атласів може бути розглянутий у вигляді алгоритму навчання з вчителем. В такому випадку ми зможемо краще утилізувати наявні анотовані зображення шляхом вивчення більш складних ознак для вибору найбільш схожих зображень атласів, особливо коли в доступі наявний великий набір даних. Прикладом такого підходу може служити алгоритм, описаний в роботі [30]. Дана робота описує алгоритм під назвою “Neighborhood Approximation Forest”, що базувався на розповсю-

дженому алгоритмі навчання з вчителем “random forest tree”. Цей алгоритм навчався складним зовнішнім ознакам, що могли відзначати найменшу відстань між цільовим зображенням та найближчими сусідами атласами. Інший алгоритм, що використовував схожі метрики подібності цільового зображення та атласу, є метод опорних векторів, що на той момент також зміг значно покращити роботи цього етапу [31].

Загалом, етап вибору зображень атласів є одним із найважливіших етапів, від якого залежить якість зображень атласів асоційованих з цільовим зображенням. А вже від якості вибраних зображень атласів залежить фінальна коректність роботи всієї системи. Для реалізації етапу вибору атласів можна використовувати прості методи, такі як вибір зображень атласів на основі метрики подібності між цільовим зображенням та анотованим зображенням із навчальної вибірки. Однак, сучасні методи для реалізації цього етапу починають використовувати алгоритми машинного навчання, які дозволяють вивчити більш складну функцію подібності між двома атласами. Вибір коректного підходу і алгоритму для вибору зображень атласів залежить від багатьох факторів, в тому числі від того, який алгоритм реєстрації зображення буде використовуватися, функціональні та не функціональні вимоги до побудови системи, такі як швидкодія, коректність роботи системи, приватність даних, можливість аудиту системи та інше. В наступному підрозділі розглянемо приклади алгоритмів реєстрації зображень, що можуть бути використанні в рамках пайплайну системи.

1.2.1.2. Етап реєстрації зображення. Другим основним етапом роботи багато-атласного методу є етап реєстрації зображення (англ. image registration). Під реєстрацією зображення розуміється задача встановлення просторової відповідності між цільовим зображенням та набором зображень атласів, вибраних на попередньому етапі. Іншими словами, етап реєстрації зображення пов’язує всі можливі значення пікселів зображень

атласів з відповідним пікселем на цільовому зображенні, для якого виконується сегментація.

Більш формально [32], відповідність набору зображень атласів \bar{A} з цільовим зображенням T може бути описано наступним чином:

$$L_i = \max_{\phi} \{sim(T, \phi(A_i) - \lambda reg(\phi))\}, \quad (1.1)$$

де операція sim є метрикою подібності між двома зображеннями, операція reg є параметром регуляризації, що контролює гнучкість процесу трансформації за допомогою константи λ , A_i - вибране зображення атлас, ϕ - просторова трансформація, за допомогою якої мітки на зображеннях атласах встановлюються у відповідність з цільовим зображенням.

Отриману мапу міток буде використано на етапі з'єднання міток відповідно до розрахованої трансформації. Більш формально форму, в якій буде використана мапа міток можна описати формулою:

$$\tilde{A} \equiv \phi(L_i) \quad (1.2)$$

Детальний огляд існуючих просторових трансформацій, що використовуються на етапі реєстрації зображень, функцій подібності та регуляризації можуть бути знайдені в [33]. Широкий огляд та порівняння методів реєстрації зображень на публічно доступних наборах даних може бути знайдено в [34].

Відповідно до роботи [33], розрізняють два види методів реєстрації зображень в залежності від вибраної моделі трансформації:

- Лінійна реєстрація
- Деформаційна реєстрація
- Ймовірнісна

Зазвичай, багато-атласні методи для кожного зображення атласу та цільового зображення розраховується одна функція реєстрації. Використання набору зображень атласів дозволяє покращити коректність та точність

роботи алгоритму реєстрації.

В підсумку, етап реєстрації зображення можна описати як етап, що використовує алгоритм побудови карти відповідності між зображеннями атласами та цільовим зображенням, на якому виконується задача сегментації. Також, створюється карта відповідності між вокселями (або пікселями у випадку двовимірного зображення) на зображеннях атласах та цільовим зображенням T . Останнім часом, найбільш популярними методами реєстрації зображення стають підходи, що базуються на моделі деформації та методах машинного навчання.

1.2.1.3. Етап злиття міток. Останнім основним етапом пайплайну багато-атласних методів є етап злиття міток (англ. label fusion). В рамках цього етапу виконується операція сегментації цільового зображення, в результаті чого буде отримана маска, яка позначатиме які анатомічні структури знаходяться на цільовому зображенні. Щоб описати цей етап більш формально [32], припустимо, що після етапу реєстрації зображення кожен із зображень атласів \tilde{A}_i та відповідних карт міток \tilde{L}_i вже зареєстровані до цільового зображення T . Результат сегментації кожного вокселя x на цільовому зображенні T (або пікселя, якщо алгоритм працює з двовимірними даними) розраховується шляхом поєднання міток на відповідній локації вокселя $L_i(x)$ на зображенні атласі. Самі перші алгоритми, що використовувалися для поєднання міток називалися методом вибору кращого зображення атласу [35] та правило голосування більшості (англ. majority voting rule) [17, 36]. В результаті, процес злиття міток та отримання міток на цільовому зображенні можна описати наступною формулою:

$$\hat{F}(x) = \arg \max_{l \in \{0,1\}} \left\{ \sum_{i \in S_T} \delta(L_i(x) = l) \right\}, \quad (1.3)$$

де δ представляє функцію, що повертає 1 якщо аргумент позитивний та 0 в іншому випадку, S_T є підмножиною індексів вибраних зображень

атласів на першому етапі.

Не дивлячись на те, що правило голосування більшості є дуже простим алгоритмом, за допомогою нього можна досягти значно вищої коректності роботи системи ніж при використанні методів з одним зображенням атласом на етапі реєстрації зображення [15]. Однак, в реальних додатках, що працюють в реальному світі, деяка кореляція в патернах помилок може існувати, що призводить до погіршення роботи системи. Тому, використовують більш надійні алгоритми поєднання міток для компенсації зазначеної вище проблеми.

Відповідно до статті [32], існують три основних категорії методів поєднання міток, що використовуються в сучасних багато-атласних методах:

- зважені методи голосування
- ймовірнісні методи
- методи машинного навчання

Розглянемо кожен із методів більш детально.

Методи зваженого голосування. Основною ідеєю методів зваженого голосування є використання вагових коефіцієнтів для оцінки важливості кожного з атласів та вибору найкращого з них. Вибір відбувається на підмножині зображень атласів, що були отримані після першого етапу. Вагові коефіцієнти відображають міру відповідності між зображенням атласом та цільовим зображенням. Така міра може визначатися як глобально так і локально, тому методи зваженого голосування розділяють на два типи [37, 38]:

- Методи, що розраховують вагові коефіцієнти використовуючи кожне зображення атлас окремо
- Методи, що розраховують вагові коефіцієнти використовуючи всі атласи шляхом мінімізації кореляції патернів помилки поміж парами зображень атласів.

В алгоритмі локального визначення міри відповідності кожне зображе-

ння атлас розглядається окремо відповідно до вагових коефіцієнтів. В такому випадку, вагові коефіцієнти відображають локальну важливість зображення атласу, та розраховують функцію відповідності між локальними патчами зображення. Зазвичай в якості метрики подібності можуть використовуватися взаємокореляційна функція (англ. cross-correlation), функція взаємної інформації (англ. mutual information), сума квадрату різниць [37] або функція емпіричного вимірювання (англ. empirical measurement)[39]. Наприклад, метрика подібності між локальними патчами цільового зображення та атласом може бути визначена як негативна експонента суми квадрату різниць [32]:

$$\omega_i(x) = \exp \left(-\gamma \sum_{y \in N_x} (T(Y) - A_i(y))^2 \right), \quad (1.4)$$

де γ є параметром нормалізації, N_x визначає значення просторового сусідства, що визначає патчі зображення центрованих в точці околу x .

В результаті, результуюча мітка буде визначатися наступною формулою:

$$\hat{F}(x) = \arg \max_l \left\{ \sum_{i \in S_T} \omega_i(x) \delta(L_i(x) = l) \right\}, \quad (1.5)$$

де $\omega_i(x)$ представляє ваговий коефіцієнт, що визначає важливість i -го зображення атласу при оцінюванні цільової мітки в локації x , δ визначає функцію, що повертає 1 якщо аргумент позитивний, та 0 в іншому випадку; S_T визначає підмножину індексів зображень атласів, що були вибрані на першому етапі пайплайну багато-атласного методу.

Існує декілька стратегій для визначення відповідності між цільовим зображенням та атласом: визначати відповідність один до одного, або визначати як один до багатьох. Стратегія відповідності один до багатьох вважається кращою, через більшу надійність такої системи та можливість отримати вищу коректність та точність роботи системи. Ще одною причиною

є те, що мітки зображення атласу розглядається в просторовому сусідстві (англ. spatial neighbourhood) [40].

Отже, для стратегії один до багатьох [32], розрахунок цільових міток може визначатися наступною формулою:

$$\hat{F}(x) = \arg \max \left\{ \sum_{i \in S_t} \sum_{y \in N_x} \omega_i(x, y) \delta(L_i(y) = l) \right\}, \quad (1.6)$$

де $\omega_i(x, y)$ є ваговим коефіцієнтом, що описує прогнозовану коректність сегментації для операції прив'язування мітки зображення атласу в точці околу $y \in N'_x$ до цільової мітки в точці околу x , N'_x є мірою просторового сусідства для пошуку потенційних зображень атласів; δ визначає функцію, що повертає 1 якщо аргумент позитивний, та 0 в іншому випадку; S_T визначає підмножину індексів зображень атласів, що були вибрані на першому етапі пайплайну багато-атласного методу.

Стратегія зваження кожного зображення атласу індивідуально працює добре коли характеристики анатомічного зображення рівномірно розподілені між всіма зображеннями атласами. Однак, при розробці додатків для реального світу це припущення не працює. На реальних зображеннях присутність деяких анатомічних характеристик та ознак може бути завищеною, що спричинить схильність системи відмічати більше занадто присутні характеристики. Для боротьби з цією проблемою використовуються методи, що працюють відразу з декількома зображеннями атласами (в англійській термінології такий підхід часто називають joint weighting strategy) [41].

Основна ідея підходу, що використовує декілька зображень атласів, полягає в мінімізації кореляції між зображеннями атласами під час етапу вибору вагових коефіцієнтів. Такий підхід дозволяє не закріплювати характеристики, що найбільш представлені в наборі даних.

Ймовірнісні підходи. Ймовірнісні підходи використовують правило

Беєсівських ймовірностей для вибору найкращих міток для цільового зображення. Для ймовірнісних підходів існує два основні сімейства алгоритмів:

- STAPLE алгоритм [42, 43], що безпосередньо оцінює показники параметрів вагових коефіцієнтів, що найкраще відповідають ймовірнісній оцінці значень цільових міток.
- Генеративні ймовірнісні моделі, що розглядають правило зваженого голосування з точки зору Баєсівського підходу [44].

Підходи на базі машинного навчання. Для вирішення етапу поєднання міток також можна використовувати методи машинного навчання. Звичайно, машинне навчання теж можна віднести до ймовірнісних методів, однак, через те що це окреме специфічне напруження, його виокремлюють в окремий клас задач. Зазвичай, для вирішення проблеми поєднання міток використовують підходи навчання з вчителем, які оцінюють взаємозв'язки між ознаками, що ми маємо на цільовому зображенні, та анатомічними ознаками, що ми маємо в зображеннях атласах [44]. Для досягнення оптимальної коректності роботи системи набір зображень атласів разом з відповідними мітками використовують для навчання системи. Ціль навчання полягає в знайденні відповідності між анатомічними структурами на вибраних зображеннях атласах, та тим що знаходиться на цільовому зображенні. Така задача вирішується подібно до задачі класифікації. Такий підхід дозволяє вивчити доволі складні функції, які допомагають достатньо швидко та без проблем визначити всі необхідні структури на цільовому зображенні. Однак, як буде показано в наступних підрозділах, ймовірнісні багато-атласні методи із поєднанням підходів глибинного навчання дозволяють замінити всі етапи непараметричних багато-атласних методів однією складною ймовірнісною моделлю та отримати значно кращі результати.

Загалом, непараметричні методи на сьогоднішній день втрачають свою популярність і починають замінюватися ймовірнісними методами на основі

глибинного навчання. Як було показано в цьому розділі, навіть у своєму класичному вигляді непараметричні методи на певних своїх етапах почали використовувати ймовірнісні моделі для виявлення більш складних патернів в даних [45].

1.2.2. Параметричні або ймовірнісні багато-атласні методи.

Як можна побачити із попередньої секції, непараметричні методи часто покладаються на певні доволі обмеженні припущення, що є специфічними для конкретного домену чи типу медичної діагностики. Такі системи одночасно працюють з декількома зображеннями в пам'яті, що негативно впливає на швидкість системи та вимагає наявності доволі великих обчислювальних потужностей. Як протипаку цьому, існує інший клас методів, що називають параметричними або ймовірнісними багато-атласними методами [46].

Методи глибинного навчання теж можна віднести до ймовірнісних багато-атласних методів, але зазвичай ці методи виділяють в окрему категорію, оскільки вони мають свою специфіку та підходи. В цій дисертаційній роботі методи глибинного навчання також виділені в окрему групу та розглядаються в окремій секції.

1.3. Методи глибинного навчання для сегментації та класифікації

Поява методів глибинного навчання мотивувалася безпорадністю класичних методів машинного навчання в класичних завданнях штучного інтелекту, такі як розпізнавання мови, розпізнавання об'єктів на зображеннях. Класичні алгоритми машинного навчання мали проблеми з експоненціальним зростанням складності узагальнення моделі до нових прикладів при роботі з даними високої розмірності. Також, були проблеми з масштабуванням таких алгоритмів до більшої кількості даних, що на практиці не

дозволяло використовувати всі наявні цифрові дані, яких з кожним роком ставало все більше [47]. А деякі алгоритми базувалися на підході локального збіжності патернів, що не працювали на дійсно складних задачах, які вимагають розуміння загального контексту [48].

Враховуючи проблеми з класичними алгоритмами машинного навчання описані вище, до популяризації та значного прогресу в застосуванні методів глибинного навчання, вважалося, що непараметричні методи працюють значно краще. Однак, такі підходи до сегментації медичного зображення працювали добре до тих пір, поки припущення про розподіл даних та гладкість функції виконувалися. Але в багатомірному просторі, в який неодмінно попадаєш коли працюєш алгоритмом машинного навчання з медичними даними, функція може бути значно складнішою та в різних локальних регіонах поводитися абсолютно по-різному. Якщо говорити специфічно про непараметричні багато-атласні методи, то навіть масштабованість до більшої кількості даних для такого класу методів є проблемою. Основна причина полягає в тому, що такий метод виконує етап реєстрації зображення використовуючи попередньо вибраний набір зображень атласів, які необхідно тримати в пам'яті. Проблеми постають тоді, коли для представлення складної функції знаходження патернів не достатньо враховувати тільки відібрані зображення на етапі селекції. Або тоді, коли таку кількість даних неможливо вмістити в пам'ять обчислювальної машини.

Ще одна проблема полягає в тому, що класичні методи машинного навчання часто покладаються на доволі жорсткі та специфічні до домену та задачі припущення, враховуючи як необхідно формувати вибірку для навчання моделі. Однак, зазвичай при побудові складної системи медичної діагностики такі припущення зробити доволі складно, через складність структури зображення та ознак які можуть використовуватися для алгоритму машинного навчання. Більше того, в задачах аналізу медичних даних завжди існує суттєва невизначеність в діагностуванні, яку часто досить важко

формалізувати і необхідна система, яка самостійно могла б знайти необхідну складну структуру чи змоделювати розподіл таких даних [49].

Описані вище проблеми можливо вирішити за допомогою методів глибинного навчання, які дозволяють моделювати складні функції та без проблем масштабуються до великої кількості даних [47].

Вперше ефективність глибинного навчання була показана на змаганні по класифікації об'єктів на зображеннях з набору даних ImageNet. Особливість цього набору даних полягає в тому, що необхідно навчити модель класифікувати більш ніж 1000 різних об'єктів, маючи в наборі близько 14 мільйонів зображень [50]. Але автори роботи [51] показали, що глибинна згорткова нейронна мережа може досягнути значно кращого результату за існуючі класичні методи, причому результат коректності (англ. accuracy) роботи моделі показував результати вищі на приблизно 8%. Такий величезний успіх привернув увагу дослідників з різних доменів і показав, що глибинні моделі показують значно вищі результати за класичні методи. Тому, з цього часу глибинне навчання почало свій активний розвиток, а науковці, що розробляють системи автоматичної діагностики, почали використовувати методи глибинного навчання і в сфері розробки систем медичної діагностики. Як показують практичні результати, методи глибинного навчання проявляють себе значно краще за класичні багато-атласні підходи в біомедичній сфері [52, 53].

Зазвичай для вирішення задач медичних додатків потрібно вирішувати не тільки задачу класифікації, а й надавати інформацію про розташування підозрілої області, тобто мітки класу повинні бути пов'язані з кожним пікселем зображення, так само як це робиться в класичних алгоритмах багато-атласних методів. Такі задачу називають задачею сегментації і вирішується спеціальним класом згорткових нейронних мереж, які будуть розглянути далі.

Однією із найбільш популярних архітектур для сегментації медичних

зображень є архітектура U-Net [54], що використовує так звану повну згорткову нейронну мережу [55] (див. рис. 1.1).

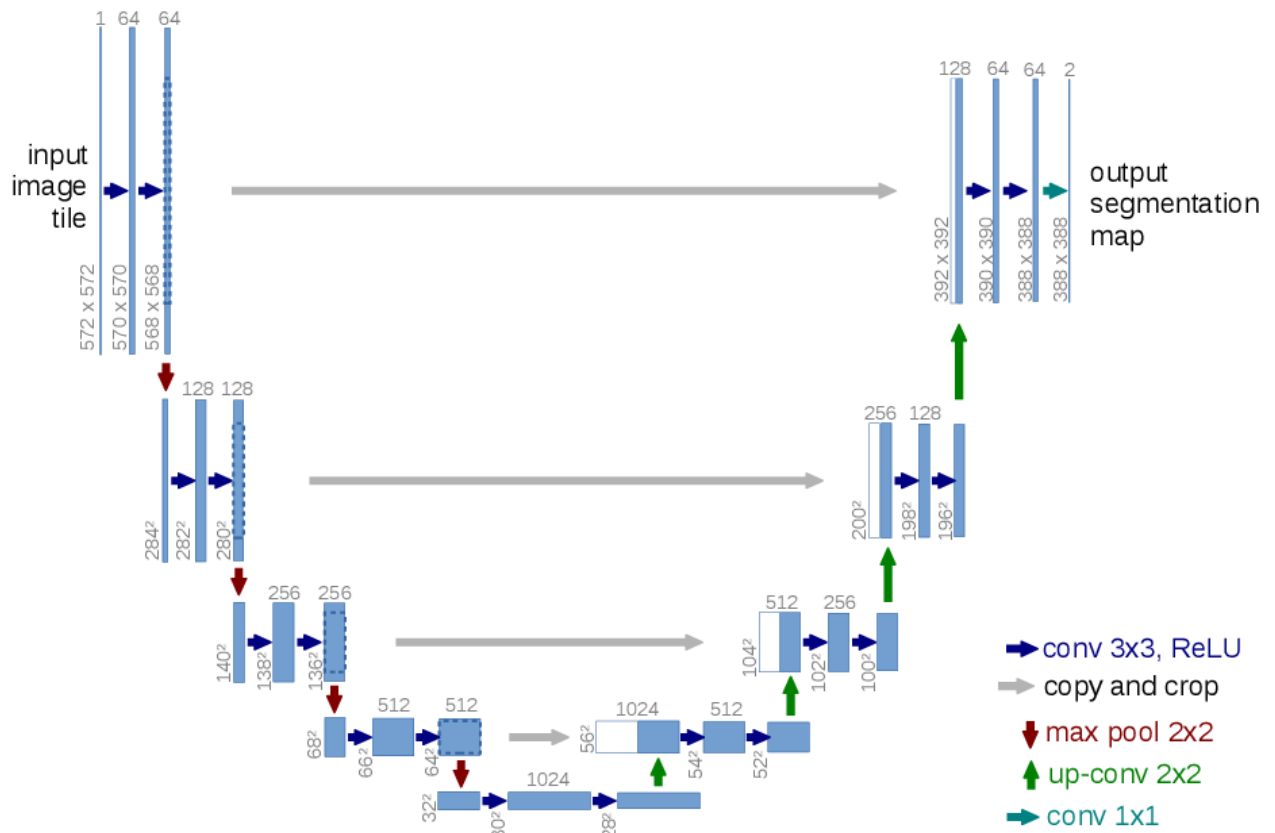


Рис. 1.1. Архітектура повної згорткової нейронної мережі U-Net, зображення взято з [54]

Архітектура згорткової нейронної мережі U-Net складається з двох частин [54]:

1. Перша частина відповідає звичайній класичній згортковій нейронній мережі, що може складатися з набору декількох операцій згортки 3×3 та операції субдискретизації за максимальним значенням (англ. max pooling) розміром 2×2 . Як функція активації зазвичай використовується ReLU. Таких блоків нейронна мережа може містити декілька. Зазвичай кількість каналів ознак згорткової нейронної мережі (англ. channels) подвоюється з кожним блоком. Приклад

першої частини згорткової нейронної мережі U-Net можна побачити на лівій частині рис. 1.1.

2. Друга частина складається з послідовних операцій оберненої згортки (англ. *transposed convolution*), що відновлюють вихідний розмір зображення, яке бути представляти собою фінальне зображення атлас існуючих структур на вхідному зображенні. Для цього використовується операція оберненої згортки, що зменшує кількість каналів ознак (англ. *channels*) вдвічі. Далі, виконуються дві звичайні операції згортки. Такі блоки повторюється стільки ж раз, скільки вони є в першій частині нейронної мережі. Ще однією особливістю архітектури U-Net є те, що кожен блок другою половиною використовує ознаки отримані на тому ж рівні в першій половині нейронної мережі. Такий підхід нівелює проблему втрати інформації при виконанні операції оберненої згортки. В кінці другої половини нейронної мережі використовується операція згортки 1×1 , що дозволяє вивести маску сегментації для вхідного зображення.

Така архітектура згорткової нейронної мережі доказала свою дієвість і дозволила досягти високих результатів при сегментації зображень, навіть якщо в наборі даних присутня доволі мала кількість даних. Так, в роботі [54] така архітектура змогла досягти високих результатів сегментації клітин на зображеннях елеронного мікроскопа.

Згорткова нейронна мережа U-Net може використовуватися не тільки для двовимірних даних. Більшість даних в медичній діагностиці мають тривимірну природу. Тому для роботи з тривимірними даними була придумана адаптація архітектури U-Net. Така архітектура згорткової нейронної мережі отримала назву V-Net [53]. Тривимірна згорткова нейронна мережа V-Net була використана для сегментації зображень магнітно-резонансної терапії (англ. *magnetic resonance imaging*), намагаючись вирішити проблему діагностики раку простати. Архітектура представленої мережі повто-

рює всі архітектурні особливості згорткової нейронної мережі U-Net, але при цьому замість двовимірних версій операцій згортки та оберненої згортки використовуються їх аналоги для тривимірного простору. Перевагою такої архітектури є врахування тривимірної природи вхідних даних.

Ще одним важливим етапом систем комп'ютерної діагностики є етап класифікації, на якому будується модель, що визначає ймовірність захворювання пацієнта на основі отриманих регіонів інтересу. Однією з популярних архітектур, що мала великий вплив на дослідження в області класифікації зображень (та ідеї цієї мережі використовують для покращення роботи нейронних мереж для сегментації) та дозволила будувати ще більш глибокі штучні нейронні мережі, була архітектура згорткової нейронної мережі під назвою ResNet [56]. Автори цієї архітектури звернули увагу на те, що в змаганнях на наборі даних ImageNet станом на 2015 рік вже намагались використовувати глибокі нейронні мережі з 16 або навіть 30 шарами. Та сама тенденція йшла і в інших складних нетривіальних задачах комп'ютерного зору. На той момент вже навчилися справлятися із проблемами градієнтного затухання та градієнтного вибуху шляхом нормалізованої ініціалізації вагових коефіцієнтів згорткової нейронної мережі та нормалізації виходів прихованих шарів. Такі підходи вже активно використовувалися в AlexNet [51]. Використовуючи їх можна було досягти гарних результатів в доволі глибоких нейронних мережах з кількістю шарів більше 10, що навчалися від початку до кінця стохастичним градієнтним спуском (англ. *stochastic gradient descent*) та алгоритмом оберненого розповсюдження помилки (англ. *backpropagation*). Однак, якщо спробувати збільшити кількість шарів згорткової нейронної мережі ще більше, скажімо до 50, чи 100, спостерігалася сильна деградація коректності роботи моделі. Така деградація метрик за спостереженням авторів відбувалася не через перенавчання, тому що помилка збільшувалася і на тренувальній вибірці. Проблеми з затуханням чи вибухом градієнтного спуску теж не було.

Тож проблема полягала в складності навчання такої моделі.

Для вирішення проблеми деградації нейронної мережі зі збільшенням кількості шарів, автори [56] запропонували ввести так звані залишкові блоки (англ. residual blocks). Автори зробили припущення та підтвердили експериментально, що алгоритму оптимізації значно простіше оптимізувати залишкове відображення, що має інформацію з попередніх етапів, ніж навчити звичайне відображення що немає таких зв'язків. Також в роботі [57] була показана структура залишкового блоку, яка була найбільш оптимальною для нейронних мереж з великою кількістю прихованих шарів. На рисунку 1.2 показано, як виглядають залишкові блоки в типовій архітектурі згорткової нейронної мережі.

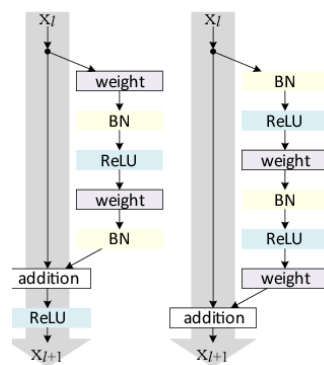


Рис. 1.2. Приклад залишкових блоків (англ. residual blocks), за допомогою яких був зроблений прорив в навчанні дуже глибоких нейронних мереж (розміром 100, 150 шарів і більше). Зображення взято з [57]

Формально, один залишковий блок в архітектурі ResNet задається наступною формулою:

$$y_l = h(x_l) + F(x_l, \Omega_L), \quad (1.7)$$

$$x_{l+1} = f(y_l), \quad (1.8)$$

де x_l та x_{l+1} є входом та виходом для l -го залишкового блоку відповідно; F - залишкова функція, що відображає нелінійність праворуч від сірих стрілок на рисунку 1.2; $h(x_l)$ - тотожне відображення, що пов'язує вхід та

вихід залишкового блоку (стрілки в сірій зоні на рис. 1.2); f - функція активації, яка зазвичай має вигляд ReLU.

Саме такий залишковий блок за словами авторів [57] дозволив навчати дуже глибокі нейронні мережі уникаючи деградації. Найбільша згорткова нейронна мережа яку тренували автори мала цілих 100 шарів і показувала покращення в коректності роботи на тестовій вибірці. Згідно результатами досліджень, залишковий блок, показаний праворуч на рисунку 1.2 є більш ефективним. Згорткову нейронну мережу з таким залишковим блоком згодом почали називати ResNetv2.

Ще одним цікавим досягненням такої архітектури було зменшення вимог до обчислювальної потужності у порівнянні з класичними архітектурами. Наприклад [56], VGG-19 вимагає швидкодію системи на рівні 19.6 мільярдів FLOPs, тоді як звичайна згорткова нейронна мережа архітектури ResNet першої версії з 34 згортковими шарами вимагає систему зі швидкодією 3.6 мільярдів FLOPs. Як можна побачити по цифрам різниця є суттєвою, тому на одному й тому ж обладнанні нейронну мережу архітектури ResNet можна навчити значно більшого розміру, чим якби слідувати більш класичній архітектурі VGG. Ця властивість є особливо корисною для побудови систем автоматичної діагностики, оскільки медичні дані зазвичай мають велику розмірність. Тому, така архітектура активно використовується і в побудові нейронних мереж для медичних додатків. Наприклад, робота [58] використовувала принципи ResNet для вирішення проблеми сегментації нирок на КТ знімку. Однак, архітектура ResNet все ще не була оптимальною з точки зору обчислювальної потужності та при збільшенні кількості шарів вимагала значно більшої кількості ресурсів. Тому згодом появились більш просунуті та ефективніші блоки згорткових нейронних мереж, що дозволили отримати ще кращі результати за рахунок більш ефективних розрахунків та меншим ростом вимог до обчислювальної потужності зі збільшенням кількості шарів. Такими архітектурами були Inception

[59, 60, 61], DenseNet [62], Xception [63].

Підходи глибинного навчання можна використати до медичних даних іншим чином, особливо до даних які представлені в тривимірному або навіть в чотирьох вимірному форматі. Проблема в тому, що медичні дані мають дуже високу роздільну здатність і чим більш сучасне та точне обладнання, тим більшу розмірність будуть мати дані. Це зумовлює проблему прокляття розмірності (англ. curse of dimensionality) для нейронних мереж, оскільки їм доводиться працювати в просторі високої розмірності. Одним із способів боротьби з цим є використання алгоритмів навчання многовиду (англ. manifold learning). Навчання таких моделей можливо завдяки припущенню, що більшість точок розподілу концентровано знаходяться в одній із областей, тобто більшість даних лежать в області менш мірного многовиду [47].

Представлене припущення можливо не є завжди правдивим або корисним, але на практиці для задач розпізнавання образів воно є як мінімум приблизно вірним [47]. Для прикладу можна взяти будь-яке медичне зображення, що може представляти знімок внутрішніх органів людини для діагностики певного захворювання. Нехай значення кожного пікселя зображення змінюється в діапазоні від $0 \dots 255$. Нехай нам потрібно згенерувати таке зображення випадковим чином щоб отримати зображення внутрішніх органів людини в цілому. Ймовірність того, що при такій випадковій генерації ми отримаємо зображення внутрішніх органів людини фактично дорівнює нулю. Для формування такого зображення нам не потрібно використовувати всі можливі варіації поєднання пікселів один з одним. Для вирішення такої задачі нам необхідно тільки дізнатися приблизні послідовності поєднання інтенсивностей пікселів, що будуть утворювати зображення внутрішніх органів пацієнта. Тобто множина необхідних варіацій поєднання пікселів медичного зображення точно менше ніж загальна кількість можливих комбінацій поєднання пікселів зображення. Виходячи з цього

можна припустити, що точно можна знайти якийсь менш мірний многовид, в рамках якого можна буде описати всі дані і при цьому спростити та пришвидшити навчання нейронної мережі.

Для навчання менш мірного многовиду класично застосовуються підходи навчання представлень (англ. representation learning). Класичними нейронними мережами в такому випадку можуть вважатися сімейство нейронних мереж автокодувальників (англ. autoencoders) [47]. Автокодувальники можуть бути використанні для знаходження такого менш мірного многовиду, що може найкращим чином описати наявні дані в просторі значно меншої розмірності. Отримані точки в такому просторі малої розмірності можна потім використовувати як вхід до стандартних архітектур нейронних мереж, наприклад тих що розглядалися вище. Також, таке зменшене представлення можна спробувати використати і в класичних непараметричних мульти-атласних алгоритмах.

Прикладом такого застосування підходів глибинного навчання можна побачити в роботі [64], де намагалися вирішити проблему детектування багатьох органів людини використовуючи доступні чотиривимірні дані пацієнтів. Автори використали навчання без вчителя для створення відображення ознак в менш вимірному просторі. Як основну модель для побудови таких ознак використали архітектуру автокодувальника під назвою "stacked sparse autoencoder". Отримані ознаки використовували як вхідні дані для більш класичного алгоритму, а саме ймовірнісних групових методів. Використовуючи такий метод автори змогли досягти гарних результатів роботи системи навіть з невеликою кількістю анотованих даних.

Автори роботи [65] розширили підхід, що був використаний в попередній роботі. Автори використали схожу архітектуру автоенкодера як частину багато-атласних методів сегментації та методів деформаційних моделей. В роботі до навченого автоенкодера були додані вихідні класифікаційні шари поверх частини кодера та навчено в стилі навчання з вчителем ви-

користовуючи наявний анотований набір даних. Результати показали, що такий підхід показує значно кращі результати чим класичні методи, що використовують ручні задані ознаки. Безперечною перевагою використання підходів на основі автокодувальників є можливість використовувати неанотовані дані.

1.4. Підсумки та постановка задачі дослідження

Системи комп'ютерної діагностики допомагають підвищити якість та швидкість діагностування пацієнтів, працюючи із складними медичними зображеннями для знаходження патологій. Стандартний пайплайн такої системи складається з етапів попередньої обробки зображення, сегментації області або органу, що аналізується, знаходження та аналізу регіонів інтересу. В 2000-х роках для виконання сегментації та виділення регіонів інтересу на медичному зображенні найбільш популярними були непараметричні багато-атласні методи сегментації, що склалися з великої кількості етапів та вимагали збереження всіх зображень-прикладів в пам'яті. Основними недоліками таких методів була їх повільна робота через використання доволі складних та дорогих операцій реєстрації зображення та об'єднання міток, високі вимоги до пам'яті, складність масштабування такої системи та невисока точність роботи.

З часом методи глибинного навчання почали активно розвиватися та показувати значно кращі результати за класичні алгоритми в різних задачах та доменах [51, 66, 67]. Тому, методи глибинного навчання почали активно використовувати на різних етапах багато-атласних методів [45], що дозволяло підвищити точність роботи та швидкодію таких систем. З часом, почали використовувати глибинні нейронні мережі, що повністю замінюють всі етапи та показували значно вищі та кращі результати, а також значно вищу швидкодію на тестовому етапі [54]. Також, точність роботи сучасних

систем комп'ютерної діагностики дозволяє додати етап класифікації, що повертає ймовірність присутності хвороби на знімку, де для класифікації також використовують методи глибинного навчання на основі отриманих регіонів інтересу.

Основний успіх методів глибинного навчання полягав у використанні згорткових нейронних мереж у задачах комп'ютерного зору. Ідея згорткових нейронних мереж не є новою та була запропонована Ян ЛеКуном ще в далекому 1998 році [68]. Однак, на той момент згорткові нейронні мережі не показували значно кращі результати через відсутність великої вибірки для навчання, достатньо потужних комп'ютерів та графічних процесорів для проведення необхідних обчислень. Основою для прогресу в 2012 році стало поєднання багатьох факторів, таких як:

- Наявність великої навчальної вибірки для навчання нейронної мережі. На той момент використовувався набір даних ImageNet LSVRC-2012, який містив близько 1.2 мільйона розмічених зображень з більш ніж 1 000 категорій у тренувальній вибірці.
- Використання функції активації ReLu (rectified linear unit) замість більш традиційних та розповсюджених функцій активацій, таких як сигмоїда та гіперболічний тангенс. Емпіричні результати експериментів показують, що функція активації ReLu підвищує ефективність навчання нейронної мережі. [69].
- Нейронна мережа навчалася з використанням графічних прискорювачів (GPU), тобто нейронна мережа була розпаралелена на двох графічних прискорювачах з доступним об'ємом пам'яті 3 гігабайти. Раніше графічні процесори не використовувалися при навчанні нейронних мереж.
- Була використана техніка нормалізації локального виходу шару нейронної мережі, щоб запобігти проблемі градієнтного вибуху або затухання.

- Використовувалися просунуті техніки регуляризації нейронної мережі для запобігання її перенавчання, а саме такі техніки як "збільшення даних" (Data augmentation) та Dropout.
- Нейронна мережа була навчена від початку до кінця використовуючи існуючу вибірку з анотованими зображеннями та підхід навчання з вчителем. Тобто мережа була навчена від початку до кінця за допомогою градієнтного спуску та алгоритму оберненого розповсюдження помилки (англ. backpropagation). Попереднє пошарове навчання нейронної мережі за допомогою навчання без вчителя не використовувалося. В оригінальній статті автори зазначають, що можливо з використанням попереднього пошарового навчання коректність (англ. accuracy) моделі була б вищою, але відмова від цього дозволила спростити та прискорити навчання за рахунок розпаралелення операцій.

Описані вище підходи навчання глибоких згорткових нейронних мереж почали використовуватися в більш пізніх дослідженнях та довели свою ефективність. На даний момент більшість сучасних задач класифікації та сегментації медичних зображень вирішуються за допомогою глибоких згорткових нейронних мереж, що навчаються від початку до кінця градієнтним спуском та алгоритмом оберненого розповсюдження помилки для розрахунку градієнтів кожного шару.

Однак, системи комп'ютерної діагностики на базі згорткових нейронних мереж в контексті побудови систем комп'ютерної діагностики для аналізу медичних зображень мають свої недоліки. Так, глибинні згорткові нейронні мережі вимагають великої кількості даних для навчання та мають проблеми з перенавчанням, особливо на невеликих вибірках. Нейронна мережа є структурою типу чорного ящика, тому роботу такої системи важко інтерпретувати, що зменшує довіру лікарів до таких систем. У розглянутих підходах та архітектурах згорткових нейронних мереж вимоги до ви-

користання пам'яті залишаються доволі високими, що ускладнює задачу навчання та використання таких систем на зображеннях з високою роздільною здатністю. Більшість медичних зображень являються собою дані високого розширення та мають тривимірну природу (наприклад зображення комп'ютерної томографії), однак існуючі підходи з двовимірними згортковими нейронними мережами не враховують це, що негативним чином впливає на точність роботи таких систем.

Враховуючи недоліки та переваги існуючих систем комп'ютерної діагностики, метою дослідження проведеного в рамках цієї дисертаційної роботи є удосконалення існуючих систем комп'ютерної діагностики на базі методів глибинного навчання, шляхом розробки системи комп'ютерної діагностики раку легенів з використанням знімків комп'ютерної томографії легенів пацієнтів. Для досягнення мети необхідно розв'язати наступні завдання:

1. Розробити або модифікувати існуючі архітектури згорткових нейронних мереж для вирішення задач сегментації та класифікації. Така архітектура має враховувати тривимірну природу КТ знімків легенів пацієнтів.
2. Розробити та використати механізми для підвищення ефективності навчання розроблених моделей.
3. Розробити власну систему комп'ютерного діагностування раку легенів на основі розроблених архітектур згорткових нейронних мереж.
4. Виконати експериментальні дослідження розроблених моделей і методів для визначення їх ефективності.

Основні матеріали, що представлені в даному розділі були опубліковані в статті [8].

РОЗДІЛ 2

ТРИВИМІРНІ НЕЙРОННІ МЕРЕЖІ ДЛЯ АВТОМАТИЧНОЇ ДІАГНОСТИКИ РАКУ ЛЕГЕНІВ

В останні роки було досягнуто суттєвого прогресу в задачах комп'ютерного зору [51], що спричинило хвилю популярності та розповсюдження підходів глибокого навчання серед науковців та інженерів. Основою прогресу стало використання двовимірних згорткових нейронних мереж з великою кількістю шарів [70]. Популяризація згорткових нейронних мереж в задачах комп'ютерного зору привернула увагу і вчених, що займаються розробкою систем автоматичної діагностики раку легенів. Як було показано в розділі 1.3, згорткові мережі почали активно використовуватися в рамках систем комп'ютерної діагностики, інколи показуючи результати коректності (англ. accuracy) роботи системи на рівні середньостатистичного кваліфікованого рентгенолога. Однак, робота з медичними зображеннями досі вважається великим викликом через складність та розмірність даних в медичній сфері, відсутності великої кількості анотованих даних. Часто, медичні дані мають тривимірну природу, що ускладнює роботу з ними. Наприклад в роботі [71] автори працюють з відео ендоскопії, а робота [72] працює з КТ зображеннями грудної клітки пацієнтів, намагаючись детектувати рак легенів.

В рамках дисертаційної роботи було проведено дослідження роботи тривимірних згорткових нейронних мереж з КТ зображеннями легенів пацієнтів (див. розділ 4.1) для діагностування раку легенів на знімку. Такі дані мають тривимірну природу, тому двовимірні згорткові нейронні мережі працюють погано (див. розділ 4.4), через те що вони втрачають інформацію про просторовий зв'язок між різними шарами знімку. Щоб мати мо-

жливість враховувати цю інформацію, було прийнято рішення використати тривимірну згортку, що не дає можливість використовувати всю доступну інформацію на КТ зображенні легенів пацієнта. В цьому розділі ми детально розглянемо операцію тривимірної згортки та різницю між тривимірною та двовимірною згортками. Також, розглянемо яким чином будується архітектура тривимірної нейронної мережі та з якими складностями зустрічаються при розробці архітектури нейронної мережі.

2.1. Згорткова нейронна мережа та її тривимірна версія

Згорткові нейронні мережі [70] у класичному їхньому вигляді представляють собою спеціалізований вид штучних нейронних мереж, що використовують операцію згортки замість добутку матриць на хоча б одному з шарів. Такі нейронні мережі використовуються для роботи з даними, що мають відому сітчасту топологію та набули популярності в задачах комп'ютерного зору при роботі з двовимірними зображеннями. Також, такі нейронні мережі можуть бути застосовані і для роботи із часовими послідовностями, що в такому випадку розглядаються як одновимірна сітка, яка формується шляхом вибору точок з часової послідовності в межах певного часового інтервалу [47].

Класичний згортковий шар складається з трьох основних етапів [70]. Перший етап представляє собою афінне перетворення у вигляді операції згортки. На другому етапі, отримані вихідні значення з афінного перетворення трансформуються через нелінійну функцію активації, яка в більшості згорткових нейронних мережах має вигляд ReLu. Інколи цей етап ще називають етапом детектора. На останньому третьому етапі виконується операція субдискретизації даних (англ. pooling) для модифікації та зменшення розмірності вхідних даних. У випадку тривимірної згорткової нейронної мережі всі етапи зберігаються у такій самій послідовності, тільки

кожна операція на кожному із етапів адаптується до роботи з тривимірними даними. Після набору послідовних згорткових шарів класична згорткова нейронна мережа використовує так званий повнозв'язний шар, що представляю собою багатошаровий перцептрон. Тобто, можна представити згорткові шари такою собі функцією витягування високорівневих ознак з вхідних даних для класичного багатошарового нейронну. Або ж, можна сказати, що згорткові нейронні мережі знаходять менш мірний многовид, що описує вихідні дані в просторі меншої розмірності. В наступних підрозділах розглянемо більш детально операцію згортки та субдискретизації для двовимірних та тривимірних згорткових нейронних мереж.

2.1.1. Операція згортки. Використання операції згортки для обробки зображень. Будь-яка згорткова нейронна мережа в своїй основі використовує математичну операцію згортки для обробки вхідного сигналу на етапі детектування. Математично, операцію згортки для неперервних величин задають наступною формулою:

$$s(t) = \int x(a)\omega(t-a)da = (x * a)(t), \quad (2.1)$$

де x - вхідні дані, які необхідно обробити за допомогою операції згортки; ω - ядро операції згортки, представляє собою матрицю вагових коефіцієнтів; t - поточний момент часу; a - час, який існують дані.

Для дискретних величин операція згортки може бути записана наступним чином:

$$s(t) = (x * a)(t) = \sum_{a=-\infty}^{\infty} x(a)\omega(t-a) \quad (2.2)$$

В термінології згорткової нейронної мережі, під x розуміють вхідні дані згорткового шару, ω позначає ядро згортки (англ. kernel). Вихідні значення операції згортки інколи можуть називати картою ознак (англ. features map) [47]. Ціль навчання згорткової нейронної мережі полягає в знаходженні таких значень вагових коефіцієнтів в ядрі операції згортки, щоб отримати

набір найбільш релевантних ознак, що будуть представляти вихідні дані в просторі меншої розмірності.

Згорткові нейронні мережі можуть працювати із двовимірними зображеннями, які у випадку чорно-білого зображення представляють собою двовимірну дискретну величину. Тому, операцію згортки для двовимірного чорно-білого зображення можна записати наступним чином:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n), \quad (2.3)$$

де I - це вхідне двовимірне зображення, K - функція ядра згортки, m та n - розмірність вхідного зображення.

Потрібно пам'ятати, що в більшості фреймворків машинного навчання, таких як tensorflow чи pytorch, операція згортки реалізується як взаємодіюча функція (англ. cross-correlation). Взаємодіюча функція виконує ті самі розрахунки що й операція згортки, але без транспонування матриці в кінці. Вважається, що в більшості випадків операція транспортування не впливає на збіжність всієї згорткової нейронної мережі [73, 74, 75]. В такому випадку записаний вище вираз згортки для двовимірних дискретних даних матиме наступний вигляд:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \quad (2.4)$$

Однак, багато вхідних даних формують не чорно-білими, а кольоровими зображеннями. Частіше всього кольорове зображення представляється в пам'яті у вигляді тривимірного масиву, кожен вимір якого характеризує інтенсивність червоного, зеленого та синього кольорів в кожному з пікселів. Представлення зображень у вигляді червоного, зеленого та синього кольорів називають кольоровою моделлю RGB. Також, згорткова нейронна мережа має цілий набір ядер у вихідному каналі, кожен з яких дивиться на свій специфічний патерн та виконується паралельно. Нехай маємо чотиривимірний масив K з елементами $K_{i,j,k,l}$, що представляють собою вагові

коефіцієнти між значенням у вихідному каналі i та вхідним каналом j та зміщенням в k рядків та l колонок між вихідними та вхідними значеннями. Також, нехай вхідні значення I складаються з елементів $I_{i,j,k}$, що представляють вхідні значення з каналу i в рядку j та колонці k . Нехай результат виконання згортки складається з значень S , що мають такий самий формат що і вхідні дані I , та який формуються шляхом застосування операції перехресної ентропії із ядром K по вхідним даним I . В такому випадку операція згортки для багатоканального входу та виходу може бути записана наступним чином [47]:

$$S_{i,j,k} = \sum_{l,m,n} I_{l,j+m-1,k+n-1} K_{i,l,m,n}, \quad (2.5)$$

де сума по l, m, n відбувається по всім значенням всіх каналів вхідного зображення.

Приклад виконання операції згортки по багатоканальному входу та одноканальним виходом можна побачити на рисунку 2.1.

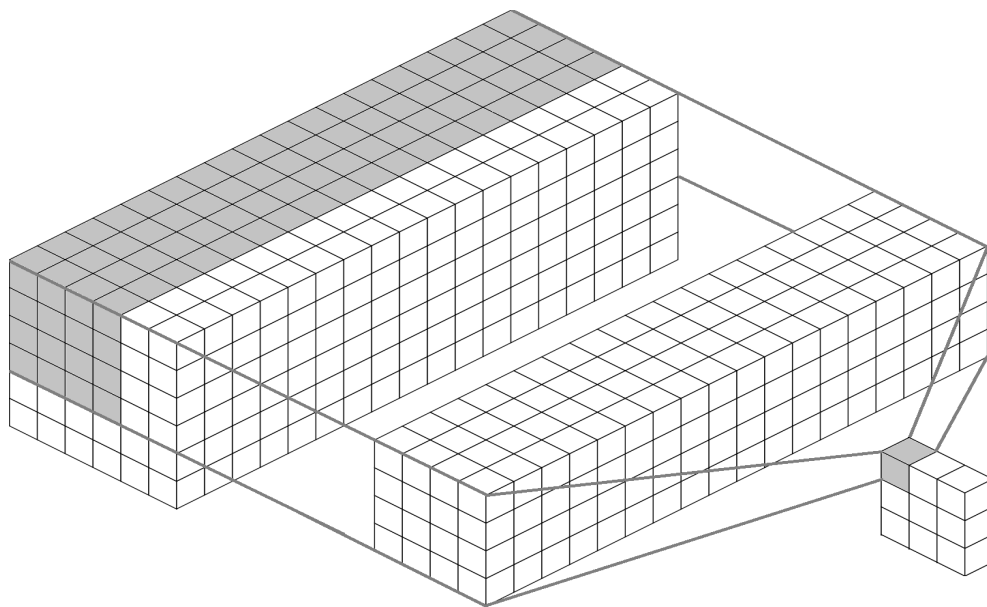


Рис. 2.1. Операція двовимірної згортки для вхідних даних з кількістю каналів більше одинці.

Як видно із зображення 2.1 та виразу 2.5, вхідні дані двовимірної згор-

ткової нейронної мережі можуть бути багатомірними. В такому випадку всі значення зі всіх каналів поелементно перемножуються із значеннями в ядрі операції згортки, та потім сумуються. Тож на виході отримується двовимірний масив з розмірністю трохи меншою за розмірність одного каналу. Така поведінка двовимірної згортки є цілком прийнятною для кольорових зображень, оскільки значення вектору інтенсивності кольорів не є залежними між собою і представляють собою одну сутність.

Однак поведінка двовимірної згортки не є прийнятною для роботи з тривимірними зображеннями, де значення по третій просторовій осі містять в собі важливу інформацію для роботи класифікатора. В КТ зображеннях легенів пацієнта завдяки третій вісі можна сформулювати уявлення про морфологію легеневого утворення. Інформація про морфологію таких утворень є важливою частиною діагностування раку легенів [76]. Застосування двовимірної операції згортки як частини згорткової нейронної мережі не дає високих результатів через те, що двовимірна згортка перетворює тривимірне вхідне зображення в двовимірні дані та втрачає всю просторову інформацію, яка була доступна за рахунок третьої осі. Щоб зберегти інформацію по всім просторовим напрямкам та мати можливість використовувати розповсюдження тривимірних даних вздовж всіх шарів згорткової нейронної мережі, необхідно замінити операцію двовимірної згортки на операцію тривимірної згортки. Як показано на рисунку 2.2, в такому разі вся просторова інформація буде збережена і ми вздовж всієї згорткової нейронної мережі працюватимем з тривимірним представленням.

Тривимірна операція згортки досягається шляхом застосування тривимірного ядра до куба вхідного тривимірного зображення. Формально, результат операції тривимірної згортки в позиції x, y, z на i -му каналі вхідного зображення можна описати наступною формулою (формула наводиться у вигляді, який використовують сучасні фреймворки машинного навчання,

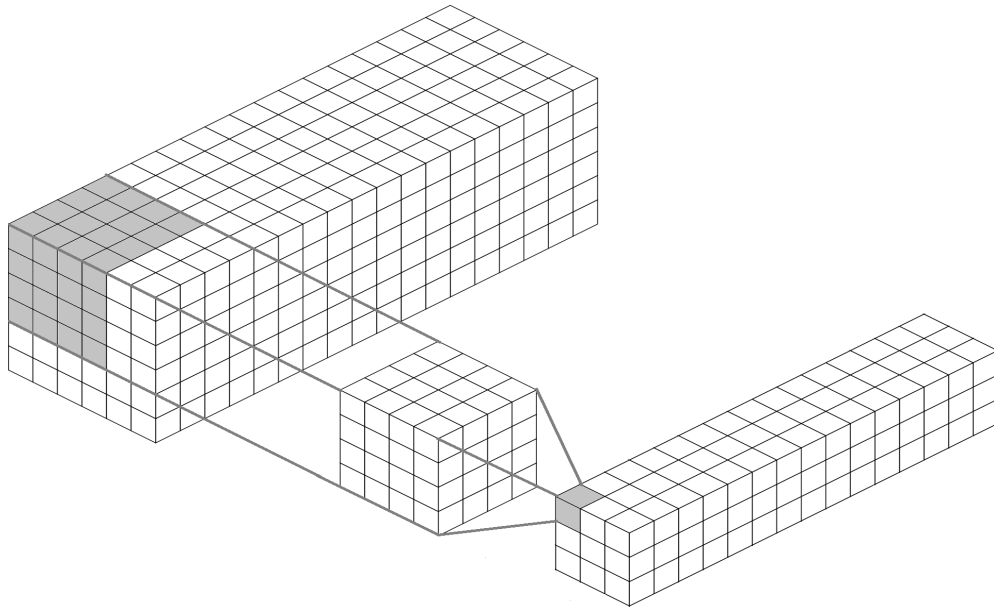


Рис. 2.2. Операція тривимірної згортки.

тобто у вигляді взаємкореляційної функції) [77]:

$$S_{i,x,y,z} = \sum_l \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} I_{l,x+p-1,y+q-1,z+r-1} K_{i,l,p,q,r}, \quad (2.6)$$

де p, q, r - координати тривимірного ядра операції згортки; K - ядро тривимірної операції згортки; I - тривимірні вхідні дані на l каналі.

Тривимірна операція згортки, подібно до двовимірної згортки, застосовує одне й теж тривимірне ядро по кожним локальним значенням вхідних даних, які мають форму кубу. Також, операція може застосовуватися з різними ваговими коефіцієнтами паралельно, формуючи набір карт ознак в кожному каналі у вигляді куба вагових коефіцієнтів.

2.1.2. Властивості тривимірної згорткової нейронної мережі.

Основні базові властивості тривимірних згорткових нейронних мереж співпадають з властивостями двовимірних мереж. Однак, такі згорткові нейронні мережі додають ще одну просторову вісь, яка дозволяє працювати та врахувати додаткові просторові залежності для тривимірних даних, та додаткові просторово-часові ознаки для діагностики відео [71]. До основних

властивостей згорткової нейронної мережі можна віднести [47]:

1. Розрідженні вагові коефіцієнти (англ. sparse weights). В традиційних нейронних мережах використовується повнозв'язні зв'язки між шарами. Однак, такий підхід вимагає поєднувати кожне значення вхідного шару з кожним значенням вихідного шару, що позначається на кількості пам'яті, яку використовує така нейронна мережа, та на швидкість роботи. Згорткові нейронні мережі не вимагають з'єднання між кожним вхідним та вихідним значеннями. Це досягається шляхом використання ядра операції згортки значно меншого розміру, ніж вхідне зображення. Для операції згортки цілком достатньо мати можливість працювати з певним локальним регіоном на зображенні. Цікаво, що більш глибокі шари згорткової нейронної мережі можуть не напряму взаємодіяти з більшою частиною вхідного зображення через отримані в певному локальному регіоні ознаки на попередніх шарах мережі. Така властивість глибоких згорткових мереж дозволяє навчити більш високорівневі абстракції на основі більш простих низькорівневих абстракцій. Для прикладу, подібну властивість можна побачити на рисунку 2.3.
2. Спільне використання вагових коефіцієнтів (англ. parameters sharing). Під спільним використанням вагових коефіцієнтів зазвичай розуміється можливість використання одних і тих же вагових коефіцієнтів в більш ніж одній функції в моделі. В традиційних нейронних мережах такого немає, оскільки матриця вагових коефіцієнтів використовується тільки одного разу при розрахунку вихідних значень для наступного шару. На противагу цьому, згорткові нейронні мережі використовують одну й ту саму матрицю ядра операції згортки для кожної з локальних частин вхідних даних, тим самим повторно використовуючи одні й ті ж самі вагові коефіцієнти для розрахунку кожного вихідного значення. Така особливість згорткових нейрон-

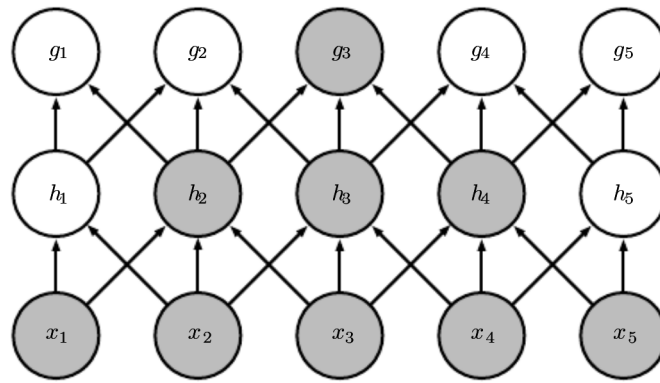


Рис. 2.3. Приклад розподілення та перевикористання навчених ознак в декількох шарах згорткової нейронної мережі. Більш глибокі шари згорткової нейронної мережі неявно використовують низькорівневу інформацію, що була отримана на більш ранніх шарах мережі. Зображення взято з [47].

них мереж означає, що замість того щоб знаходити набір вагових коефіцієнтів для кожної з локацій на вхідному зображенні, навчається тільки один такий набір. Завдяки цьому значно зменшується використання пам'яті та підвищується статистична ефективність такої нейронної мережі. Також в переваги можна додати й те, що таким чином проблема перенавчання трохи зменшується.

3. Еквіваріантність представлення (англ. equivariant representation). Під властивістю еквіваріантності в згорткових нейронних мережах розуміють те, що при певній зміні вхідних значень вихідні значення будуть змінені відповідним чином, тобто симетрично. Тобто, функцію $f(x)$ називають еквіваріантною до функції g якщо виконується рівність $f(g(x)) = g(f(x))$. З точки зору обробки часових рядів, така властивість означає, що операція згортки повертає певного виду хронологію, що показує в який момент та які ознаки були отримані на вході. Якщо одна й та сама подія буде проявлятися пізніше в часі, то вона матиме точно таке саме представлення, що мало раніше. Для тривимірних медичних даних, такі як КТ зображення

легенів пацієнта, тривимірна згорткова операція буде показувати де певна ознака знаходиться на вхідному зображенні. Також, якщо певний об'єкт був значно збільшений у розмірах (наприклад, нове КТ зображення пацієнта відображає збільшення пухлини в легенях пацієнта), то такий об'єкт буде представлено відповідним чином і після виконання операції згортки.

Описанні вище операції згортки показують, чому згорткові нейронні мережі зіграли настільки важливу роль в історії глибинного навчання. Згорткові нейронні мережі були першими робочими глибинними нейронними мережами, що були навчені з використанням алгоритму розповсюдження помилки, тоді як загальні нейронні мережі прямого розповсюдження розглядалися як неуспішні та не показували гарних результатів [47]. Такі результати були досягнуті в тому числі через описані властивості операції згортки, що дозволило будувати більш обчислювано та статистично ефективні моделі, які значно легше навчити у порівнянні з класичним багатошаровим перцептроном. Ще однією особливістю таких мереж є те, що навіть враховуючи відсутність повнозв'язних зв'язків на кожному з шарів, така нейронна мережа може вивчити високорівневі ознаки базуючись на елементарних низькорівневих ознаках, отриманих на попередніх більш ранніх шарах. Наприклад, на зображенні 2.4 показується ознаки, які шукає навчена на ImageNet згорткова нейронна мережа на різних рівнях архітектури.

На більш низькорівневих шарах мережа дивиться на певні елементарні прості структури та контури, тоді як вже на більш високорівневих шарах вже промальовуються цілком впізнавані об'єкти, які може класифікувати дана нейронна мережа. Таким чином, згорткові шари нейронної мережі представляють такий собі витягач високорівневих ознак для класичного багатошарового перцептрона. Іншими словами, представляє вхідні дані у вигляді многовиду меншої розмірності.

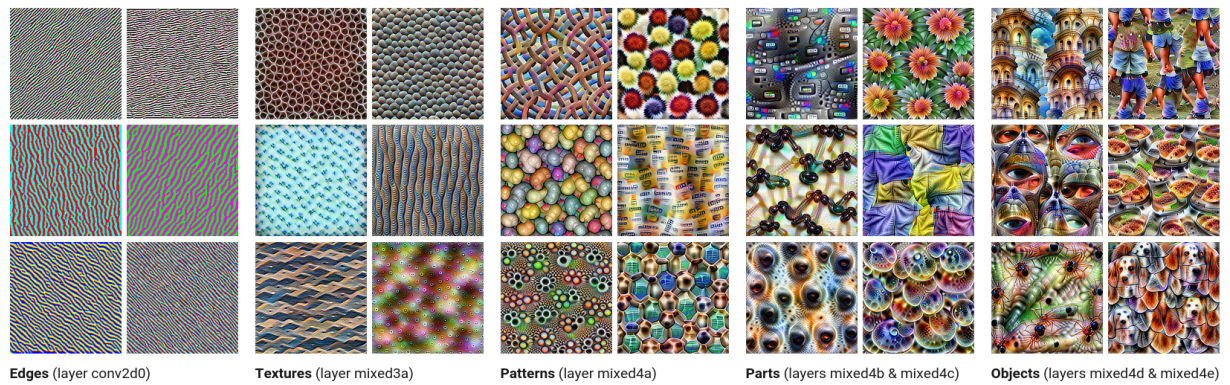


Рис. 2.4. Згорткова нейронна мережа навчається все більш високорівневим ознакам та патернам на більш глибоких шарах. Зображення показує візуалізацію ознак GoogLeNet [59], навченої на ImageNet. Зображення взято з [78].

2.1.3. Операція субдискретизації. Її властивості та використання. Операція субдискретизації (англ. pooling) є функцією, яка використовується в кінці шару класичної згорткової нейронної мережі. Задача такої функції полягає в зменшенні розмірності вхідної карти ознак, отриманої після застосування нелінійної функції активації. Це досягається шляхом збору та агрегації сумарних статистик найближчих сусідніх значень. Статистика може використовуватися різна, вибір залежить від конкретного домену та може підбиратися експериментальним чином. Наприклад, робота що вперше описала згорткові нейронні мережі в поточному їхньому вигляді [68], використовувала середнє значення найближчих значень вихідних карт ознак. На даний момент в сучасних архітектурах згорткових нейронних мереж найбільш популярною є функція субдискретизації за максимальним значенням (англ. max pooling) [79]. Операція субдискретизації за максимальним значенням розраховує максимальне значення в рамках певного прямокутника на карті ознак, що була отримана після застосування нелінійної функції активації на етапі детектора. Така функція дискретизації показує себе добре на більшості даних.

Операція субдискретизації грає важливу роль в досягненні властивості інваріантності до невеликих змін в згорткових нейронних мережах. Під інваріантністю моделі розуміють властивість системи не змінювати свій стан при невеликих змінах у вхідних значеннях. Властивість локальної інваріантності є важливою характеристикою в моделях, де пріоритетніше перевірити наявність певної ознаки, ніж її точна локація. Наприклад, при діагностиці раку легенів нам більш важливо сам факт наявності пухлини, ніж її наявність. Конкретна локація пухлини може бути приблизною та необхідна тільки для надання рентгенологу додаткової інформації про те, що в певному регіоні знімку знаходиться пухлина. Можливість не змінності локації навіть при невеликих модифікації цього регіону також є цілком припустимою [47].

Операція субдискретизації підсумовує інформацію всіх ближніх точок на карті ознак, отриману після етапу детектування згорткового шару. Така властивість субдискретизації дозволяє використовувати субдискретизацію значно менше разів, чим наприклад етап детектування. Також, розмірність даних зменшується при проведенні такої операції, що є важливим чинником при роботі із зображеннями високої роздільності, з якими в основному необхідно працювати в рамках медичних додатків. Такі характеристики і результати роботи субдискретизації допомагають покращити обчислювальну ефективність згорткової нейронної мережі. Однак, в рамках операції субдискретизації втрачається доволі велика кількість інформації, яка могла би бути корисною для покращення коректності роботи згорткової нейронної мережі. Тому, сучасні архітектури згорткових нейронних мереж намагаються використовувати як можна меншу кількість шарів субдискретизації або взагалі відмовитися від них та використовувати операцію глобальної субдискретизації.

Операція глобальної субдискретизації полягає у виконанні функції субдискретизації по всій карті ознак замість використання локального па-

тчу зображення [80]. Потім, отриманий вектор подається на вхід softmax функції для визначення ймовірності належності вхідних даних до одного з класів. Таким способом можливо уникнути використання повнозв'язного шару згорткової нейронної мережі, що вирішить проблему перенавчання на цьому шарі, оскільки на етапі глобальної субдискретизації немає параметрів, які могли б бути оновлені на етапі навчання. Також, в порівнянні з багатозв'язним шаром, глобальна субдискретизація є більш природною структурою для згорткової нейронної мережі. Операція глобальної субдискретизації встановлює відповідність між картами ознак і категоріями, які ми маємо віднайти на зображенні.

2.1.4. Навчання та регуляризація тривимірних згорткових нейронних мереж. Навчання тривимірних згорткових нейронних мереж відбувається таким же чином як і для двовимірних згорткових нейронних мереж — за допомогою звичайних технік навчання з вчителем без попереднього пошарового навчання шарів згорткової нейронної мережі. Для навчання використовується навчальна вибірка, екземпляр кожної з якої представляє собою інформацію про вхідне тривимірне зображення та асоційовану мітку з ним. Тривимірна згорткова нейронна мережа є повністю диференційованою моделлю та навчається за допомогою алгоритму градієнтного спуску (англ. gradient descent) та зворотнього розповсюдження помилки (англ. backpropagation) для розрахунку значень градієнтів для кожного з шарів. Основна відмінність та складність тривимірних згорткових нейронних мереж полягає у тому, що така нейронна мережа має значно більшу кількість параметрів чим аналогічного розміру двовимірна згорткова нейронна мережа. Вимоги до обчислювального кластеру та пам'яті ростуть кубічно із збільшенням роздільної здатності вхідного зображення. Наприклад для КТ зображень легенів пацієнта, що використовувалися в цій роботі (див. розділ 4.1), потрібно значно зменшувати роздільну зда-

тність КТ знімку, щоб мати можливість вмістити хоча б один тренувальний приклад в пам'ять комп'ютера з більш-менш глибинною тривимірною згортковою нейронною мережею.

Присутність значно більшої кількості параметрів робить тривимірну згорткову нейронну мережу складнішою для навчання через те, що така мережа ще більше схильна до перенавчання, особливо коли використовується невеликий за розміром набір даних. Також, проблеми градієнтного затухання та вибуху присутні так само і при навчанні тривимірної нейронної мережі, що вимагає використання додаткових технік для покращення навчання.

Проблема затухання градієнту полягає в тому, що градієнти можуть ставати все меншими і меншими з тим як алгоритм оберненого розповсюдження помилки проходить від високорівневих шарів до більш низькорівневих. В результаті такої поведінки градієнтний спуск при оновленні вагових коефіцієнтів фактично не змінює їх або взагалі залишає незмінними, що призводить до незбіжності нейронної мережі. На противагу постійному зменшенню градієнту може відбуватися і обернений процес, коли з кожним шаром градієнт може ставати все більшим та більшим, поки вагові коефіцієнти певних шарів не будуть оновлені на величезне значення, що призведе до розходження алгоритму. Таку проблему називають проблемою градієнтного вибуху. В перше певні причини такої нестабільної поведінки градієнтів під час навчання було описано в роботі [81], де пропонувалося при використанні сигмоїдної функції активації використовувати спеціальну техніку ініціалізації вагових коефіцієнтів. На даний момент для зменшення проблеми нестабільності градієнтів рекомендується використовувати функцію активації Relu (або будь-яку її модифікацію) та спеціальну стратегію ініціалізації вагових коефіцієнтів, яку називають He initialization [82]. Однак такі підходи повністю не вирішують проблему і в сучасних згорткових нейронних мережах, тому використовуються більш просунуті техніки

для нормалізації виходів з проміжних шарів згорткової нейронної мережі, які в цілому застосовуються і для тривимірних згорткових нейронних мереж також.

Для підвищення стабільності градієнтів під час навчання нейронної мережі в 2015 році був запропонований метод нормалізації по міні-групам. (англ. batch normalization) [83]. Така техніка полягає у використанні операції нормалізації перед або після функції активації на кожному прихованому шарі згорткової нейронної мережі. Мета нормалізації по міні-групам полягає у центруванні середнього значення та нормалізації вхідних даних, після чого вихідні значення масштабуються та зміщуються. Масштабування та зміщення відбувається за рахунок двох нових векторів, оптимальні значення яких знаходяться протягом навчання нейронної мережі. Тобто, нормалізація по міні-групам дає можливість мережі знайти найбільш оптимальні параметри масштабування та середнього значення для вхідних значень кожного з шарів. Під час навчання операція нормалізації по міні-групам прогнозує середнє значення та дисперсію кожного вихідного значення. Така статистика вираховується по міні групам, на які розділена навчальна вибірка. Метод також можна застосовувати і при навчанні відразу на всіх даних, але зазвичай це не є практично, якщо розмір тренувального набору та розмірність даних є дуже великими.

Для прогнозування дисперсії та середнього значення виходу, метод нормалізації за міні-групами розраховує середнє значення та середнє квадратичне відхилення по міні-групах доступних на етапі навчання. Розрахунки відбувається за наступними виразами:

$$\mu_B = \frac{1}{m_B} \sum_{i=1}^{m_B} x^{(i)}, \quad (2.7)$$

$$\sigma_B^2 = \frac{1}{m_B} \sum_{i=1}^{m_B} \left(x^{(i)} - \mu_B \right)^2, \quad (2.8)$$

$$\hat{x}^{(i)} = \frac{x^{(i)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad (2.9)$$

$$z^{(i)} = \gamma \otimes \hat{x}^{(i)} + \beta, \quad (2.10)$$

де μ_B вектор середніх значень, розрахований для всіх екземплярів в міні-групі B (одне середнє значення для кожного екземпляру); σ_B представляє вектор середньо-квадратичного відхилення, так само розрахований для кожного з екземплярів в міні-групі B ; m_B представляє кількість прикладів в міні-групі B ; $\hat{x}^{(i)}$ представляє вектор нормалізованих та відцентрованих за середнім значенням екземплярів в міні-групі B ; γ - вектор параметрів масштабування для кожного із вхідних екземплярів в міні-групі B . Вектор задається для кожного шару мережі в якому використовується нормування за міні-групою; \otimes - операція поелементного перемноження матриць; β - вектор вихідних значень зміщень для кожного із вхідних екземплярів в міні-групі B ; ϵ - дуже мале число, що протидіє діленню на нуль; $z^{(i)}$ - вихідні значення отриманні після застосування нормування за міні-групою. Вихідне значення представляє собою масштабовану та зміщену версію значень вхідних екземплярів.

В загальному випадку, нормалізація за міні-групою може бути застосована до будь-яких вхідних значень в нейронній мережі та в оригінальній роботі застосовувалося до операції згортки в нейронній мережі та до застосування нелінійної функції активації. Застосування нормалізації за міні-групою в загальному випадку може бути записано наступним чином [83]:

$$out = \sigma(BN(WI)), \quad (2.11)$$

де BN трансформація застосовується індивідуально до кожного просторової вісі WI , з індивідуальною парою параметрів $\gamma^{(k)}$ та $\beta^{(k)}$.

В рівнянні 2.11 відсутній параметр зміщення b , тому що його вплив нівелюється при нормуванні за міні-групою. При застосуванні до операції

згортки, нормалізація застосовується спільно до всіх вхідних значень та у всіх локаціях двовимірних даних, що знаходяться в міні-групі. Алгоритм застосовується таким чином для гарантування того, що різні елементи в різних локаціях карти ознак нормалізувалися однаково і тим самим забезпечується збереження еквіваріантності згортки. Такий підхід добре зарекомендував себе та допоміг досягти значного прогресу в вирішенні задачі класифікації на наборі даних ImageNet. Емпіричні експерименти показали, що така нормалізація змогла значно зменшити проблему градієнтного затухання та вибуху при навчанні глибоких нейронних мереж. Як бонус, нормалізація за міні-групою має невеликий ефект регуляризації та робить нейронну мережу менш чутливою до способу ініціалізації вагових коефіцієнтів. Цей підхід став де факто стандартом при навчанні як двовимірних згорткових нейронних мереж, так і тривимірних згорткових нейронних мереж [72].

Використання нормалізації по міні-групам допомагає пришвидшити збіжність глибоких нейронних мереж, але в деяких випадках цього є недостатньо. Для побудови систем медичної діагностики зазвичай доступно доволі обмежена кількість даних, тому глибокі нейронні мережі повинні могли навчатися навіть на зовсім невеликих вибірках. Тому, для пришвидшення навчання мережі та можливості отримати достатній рівень коректності роботи на малій вибірці використовують техніку навчання перенесенням (англ. transfer learning).

Суть методу навчання за перенесенням полягає в тому, що нейронна мережа попередньо навчається на існуючому великому наборі даних на схожій задачі. Наприклад, якщо необхідно вирішувати задачу класифікації по двовимірним зображенням, часто попередньо навчають глибоку нейронну мережу на наборі даних ImageNet. Далі, навчена нейронна мережа заморожує всі згорткові шари, або тільки низькорівневі, та замінює повнозв'язний шар на повністю новий з заново ініціалізованими ваговими коефіцієнтами.

В такому режимі нейронна мережа навчається на доступному наборі даних для задачі, що вирішується. Заморожені шари згорткової нейронної мережі не приймають участі при навчанні. Вибір заморожувати всі згорткові шари, або тільки їх частину, залежить від того, як багато даних доступно для навчання у вибірці, що використовується для нової задачі. Такий спосіб добре зарекомендував себе та пришвидшує навчання глибоких згорткових нейронних мереж. Це обумовлюється властивістю згорткових мереж навчатися ознакам в ієрархічному порядку (див. рис. 2.4). Для більшості задач, що працюють із зображеннями, низькорівневі ознаки є однакові, а ознаки високого рівня відрізняються в залежності від типу об'єктів, що класифікуються мережею. Для не глибоких нейронних мереж такий підхід працювати не буде.

Для тривимірних згорткових нейронних мереж знайти попередньо навчену мережу на схожих задачах неможливо. Причиною цьому є те, що вони є не надто розповсюджені, тож у відкритому доступі є невелика кількість попередньо навчених нейронних мереж. Наприклад, існує попередньо навчена тривимірна згорткова нейронна мережа C3D [84] та I3D [85]. Обидві моделі навчені на наборах даних для задачі класифікації відео. Такі попередньо навчені моделі будуть корисними для медичних додатків, що також працюють з відео, наприклад з відео ендоскопії [71]. Однак такі нейронні мережі не дадуть результату якщо повторно використати їх вагові коефіцієнти для задач класифікації тривимірних зображень, таких як класифікації раку легенів на КТ знімках пацієнту. Причина цьому є те, що відео та тривимірні дані мають різну природу і тому мають різні низькорівневі ознаки. В такому випадку нейронна мережа навчається з нуля.

Нейронна мережа з великою кількістю параметрів та глибокою структурою має тенденцію перенавчатися, особливо на малій вибірці даних. Один із способів боротьби з цією проблемою є збільшення кількості екземплярів в наборі даних. Однак, в медичній сфері часто немає можливості

отримати більше даних у зв'язку з складністю проведення діагностики та доступу до приватних даних пацієнтів. Тому, одним із способів “віртуально” збільшити набір даних є використання техніки збільшення даних (англ. data augmentation). Суть цього підходу полягає у тому, що існуючі дані певним чином випадково трансформуються під час навчання моделі та використовуються поряд з іншими реальними екземплярами в тренувальній вибірці. Існує три основні підходи для використання техніки збільшення даних:

1. Генерація розширеного набору та збільшення даних через застосування набору заданих трансформацій. В цьому випадку фізично збільшується кількість даних, але по факту нові згенеровані дані є просто певною трансформацією існуючих зображень. На кожній епосі навчання нейронна мережа має інформацію про всі дані, як і оригінальні, так і трансформовані. Цей підхід використовується не дуже часто, через те що він не допомагає значно покращити узагальненість моделі. Причина цього полягає в тому, що нейронна мережа може помітити те, що деякі дані є просто трансформацією інших та не враховувати їх.
2. Змінювати та трансформувати навчальну вибірку прямо під час навчання. Трансформовані дані в цьому випадку будуть замінювати оригінальні. Цей підхід збільшення даних є найбільш популярним і розповсюдженим, часто використовується за замовчуванням в сучасних фреймворках машинного навчання. Такий підхід веде до кращої узагальненості моделі, хоч і може зменшувати коректність роботи моделі на тренувальній вибірці.
3. Поєднує два попередніх підходи. Попередня генерація, трансформація та збереження цих даних в існуючу тренувальну вибірку з оригінальними даними. Також, під час навчання деякі оригінальні дані можуть бути замінені випадковою трансформацією. Такий

підхід теж часто використовується, особливо для задач де важко і дуже дорого збирати реальну інформацію. Наприклад, для генерації додаткових даних можуть використовуватися симулятори, як це може робитися для задач автономного керування автомобіля.

Види трансформацій, що можуть використовуватися, залежать від виду та природи даних, домену та задачі яку ми вирішуємо. Використання підходу збільшення даних в даній роботі для КТ зображень легенів пацієнтів можна знайти в розділі 4.3.

Ще одним способом боротьби з перенавчанням згорткових нейронних мереж, як двовимірних так і тривимірних, є використання методів регуляризації, що зменшують складність моделі шляхом додавання штрафного значення $\Omega(W)$ до функції втрат L :

$$L(W, x, y) = L(W, x, y) + \lambda \Omega(W), \quad (2.12)$$

де L - функція втрат; W - вагові коефіцієнти моделі; x - вхідні значення, що передавалися під час навчання; y - очікувані результати; λ - гіперпараметр, що дозволяє регулювати силу штрафу, що застосовується до моделі; $\Omega(W)$ - функція, що застосовується для розрахунку штрафу для моделі.

Найбільш розповсюдженими значеннями $\Omega(W)$ є l^2 та l^1 норми, тому такі види називають l^2 та l^1 регуляризацією відповідно. Значення регуляризації додається до функції втрат тільки під час процесу навчання. l^2 регуляризація допомагає нейронній мережі краще узагальнювати дані шляхом дотримання як можна менших значень вагових коефіцієнтів. l^2 регуляризація може бути записана наступним чином:

$$\Omega(W) = \frac{1}{2} \|W\|_2^2 = \frac{1}{2} \sum_{i=1}^n W_i^2 = \frac{1}{2} W W^T \quad (2.13)$$

Особливістю l^1 регуляризації є зменшення впливу вагових коефіцієнтів для тих ознак, що є найменш важливими для прийняття фінального

рішення моделлю. l^1 регуляризація може бути записана наступним чином:

$$\Omega(W) = \|W\|_1 = \sum_i |W_i| \quad (2.14)$$

Ще одним популярним методом регуляризації, що є специфічним саме для роботи з глибинними нейронними мережами і може бути адаптований до тривимірної згорткової нейронної мережі, є метод дропаут або метод виключення [86, 87]. Даний метод довів свою ефективність в задачах регуляризації глибинних нейронних мереж покращуючи роботу навіть тих мереж, що вже показують коректність роботи на рівні 95%. В основі дропаута лежить дуже проста проста ідея - під час навчання нейронної мережі кожен нейрон може бути тимчасово виключений та проігнорований протягом всієї тренувальної епохи [88]. Наступної епохи раніше виключені нейрони можуть стати активними, тоді як інші нейронні будуть вибрані для ігнорування. Виключеними можуть бути будь-які нейрони окрім вихідних. Нейрони виключаються випадковим чином. Кількість нейронів, що будуть виключатися під час навчання задається гіперпараметром p безпосередньо перед навчанням, а оптимальне значення підбирається під час пошуку гіперпараметрів мережі. Для згорткових нейронних мереж зазвичай рівень нейронів, що ігноруються на тренувальному етапі вибирається рівним 40-50%, але в залежності від задачі і набору даних значення можуть варіюватися. Після навчання нейронна мережа не використовує дропаут та працює зі всіма нейронами. Але, вагові коефіцієнти вхідних значень масштабуються на значення $1 - p$ для нормалізації очікуваного виходу.

2.2. Використання тривимірної згорткової нейронної мережі для діагностування раку легенів

Тривимірні згорткові нейронні мережі є досить популярними та використовуються як і для обробки відео, так і для даних, що мають триви-

мірну природу. Для задачі діагностування раку легенів, використовуються зображення комп'ютерної томографії (КТ) грудної клітки людини (див. розділ 4.1), що отримується таким чином, щоб зображення містило знімок легенів пацієнта. По суті, КТ зображення представляє собою тривимірне відображення легенів пацієнта, разом з іншими частинами тіла людини, такі як судини чи кістки. Для отримання тривимірного зображення необхідно поєднати разом всі шари КТ, що були отримані під час сканування грудної клітки людини. Таке тривимірне зображення можна використовувати як вхід для тривимірної згорткової нейронної мережі.

Зазвичай тривимірне зображення, отримане за допомогою комп'ютерної томографії, містить в собі інформацію не тільки про тканини людських легенів. Таке зображення також містить відображення кісткової тканини, судин які знаходяться в грудній клітці людини. Тому, перед використанням згорткової нейронної мережі зазвичай виконуються кроки по попередній обробці зображення, які видаляють відображення, що не відносяться до легенів. Кроки попередньої обробки даних, що використовувалися в даній роботі детально описуються в розділі 4.2. Така попередня обробка допомагає пришвидшити та спростити навчання нейронної мережі, дозволяючи отримати кращу коректність (англ. accuracy) роботи нейронної мережі. Якщо для навчання доступна велика кількість даних для тренування тривимірної нейронної мережі та доступна достатньо велика обчислювальна потужність для побудови великої нейронної мережі, то можливо навчити мережу без попередньої обробки. В такий ситуації, нейронна мережа може знайти та навчитися більш складним патернам.

Для роботи із тривимірним зображенням легенів пацієнта, отриманим за допомогою КТ та попередньої обробки, можна використовувати різні стратегії та види тривимірних згорткових нейронних мереж. Найбільш проста стратегія складається всього з одного етапу, в рамках якого вибрана

тривимірна згорткова нейронна мережа виконує задачу класифікації присутності раку на зображенні. В такому пайплайні можна використати такі згорткові нейронні мережі як C3D [89] або 3D DenseNet [90]. В рамках дисертаційної роботи були навчені ці нейронні мережі на задачі класифікації раку легенів та було отримано результат коректності роботи цих нейронних мереж на рівні 63% та 70% відповідно. Самі нейронні мережі навчалися на тривимірному зображенні, що було зменшено в цілях економії пам'яті та можливості помістити тривимірні згорткові нейронні мережі в пам'яті графічного процесору. Прикладом навченої 3D DenseNet архітектури моделі можна побачити на рисунку 2.5. Більш детальна інформація про навчання цих нейронних мереж може бути знайдена в розділі 4.6.

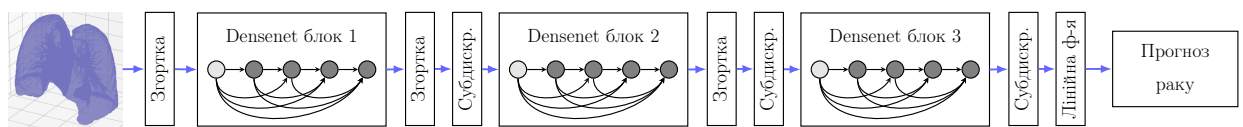


Рис. 2.5. Тривимірна версія архітектури DenseNet.

З отриманих результатів стає очевидним, що на наявному наборі даних одноетапна тривимірна згорткова нейронна мережа не показує дуже гарних результатів. Однією з причин таких невисоких результатів може бути в тому, що при подачі на вхід нейронної мережі подавалася значно зменшене в розмірах тривимірне зображення. При збільшенні розмірності вхідного зображення нейронній мережі буде доступно більше інформації, яка могла бути нівельована в зображенні меншої розмірності. Така додаткова інформація може допомогти моделі навчитися більш складним ознакам, що підвищило б коректність роботи моделі. Також, причина такої низької коректності роботи навчених тривимірних згорткових нейронних мереж може полягати в тому, що наявна тренувальна вибірка недостатньо велика для того, щоб побудовані мережі могли вивчити більш складні ознаки та враховувати більше інформації про природу злоякісних утворень в зображенні

легенів. Для того, щоб побудована система могла використовувати більше інформації та показувала кращі результати, потрібно розділити задачу на два більш простих етапи. Як показано в роботах [72] та [91], розділення системи діагностики раку на два більш прості етапи дає значне підвищення коректності діагностики та навіть дозволяє отримати коректність роботи системи на рівні середньостатистичного кваліфікованого рентгенолога.

Головною ідеєю двоетапного підходу побудови систем автоматичного діагностування раку полягає в тому, що задача діагностування раку легенів може бути розділена на декілька підзадач: сегментація та класифікація (див. рис. 2.6).

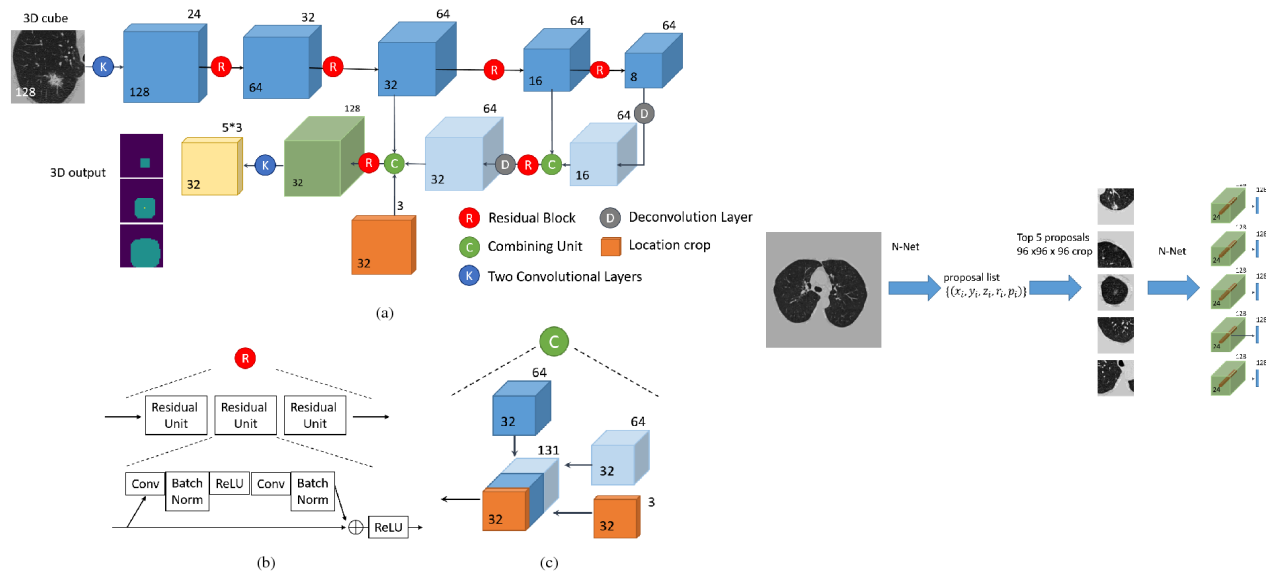


Рис. 2.6. Зліва: Етап сегментації. На правій частині: етап класифікації. Побудована двоетапна система в [72]. Зображення взяті з оригінальної статті.

На першому етапі вирішується задача сегментації, в рамках якої нейронна мережа буде виділяти підозрілі регіони із легеневими утвореннями, що ймовірно містять пухлину. Далі, система вибирає m регіонів, які найбільш ймовірно містять ракову пухлину та передає нейронній мережі, що буде виконувати задачу класифікації. На другому етапі виконується задача класифікації, яка робить фінальний висновок про те, чи присутній рак на

КТ зображені пацієнту чи ні, а також, в якому конкретно із вибраних регіонів. Одна із проблем побудови цього етапу полягає в тому, що нейронна мережа отримує на вхід п'ять різних зображень, на яких потенційно може бути злоякісне утворення, але в той же час, ми маємо тільки одну мітку про присутність раку для всього тривимірного зображення. Тому, для навчання нейронної мережі в цьому місці застосовується припущення навчання за набором зразків (англ. *multiple instance learning*). Припущення говорить про те, що якщо у всьому наборі зразків присутній хоча б один позитивний приклад, то тоді весь набір даних вважається позитивним. Весь набір зразків вважається негативним, якщо весь набір є негативним (напр. якщо мітка в тренувальному наборі даних говорить про те, що на даному знімку КТ відсутні рак легенів).

Однією з переваг у двоетапному підході для діагностування раку легенів є можливість використати одну із найбільш розповсюджених згорткових нейронних мереж для сегментації зображень та на виході отримати маленькі тривимірні патчі із злоякісними утвореннями. Завдяки цьому, на етапі класифікації є можливість використовувати більш складні та глибокі моделі, оскільки для них доступно більше пам'яті на графічному процесорі. Також, двоетапний підхід допомагає навчити систему детекції раку легенів на меншій кількості тренувальної вибірки.

Для задачі сегментації вхідного тривимірного зображення, використовуються такі згорткові нейронні мережі як V-Net [53], Faster-RCNN [92] та Mask-RCNN [93], адаптовані до тривимірних вхідних даних шляхом заміни двовимірної згортки на тривимірну згортку. Наприклад, в роботі [72] для задачі сегментації була адаптована нейронна мережа Faster-RCNN.

Зазвичай, на етапі класифікації використовується якась проста повнозв'язна згорткова нейронна мережа. Ця нейронна мережа вирішує задачу бінарної класифікації, та повертає інформацію про те, чи дана область містить злоякісне утворення. Нейронна мережа для класифікації може по-

вторно використовувати ті самі вагові коефіцієнти, що і згорткова нейронна мережа на попередньому етапі. Завдяки цьому прискорюється процес навчання та підвищується коректність роботи нейронних мереж на обох етапах. Згідно з роботами [72] та [91] двоетапні системи діагностування раку легенів дозволяють досягнути високої точності роботи системи та могли досягнути коректності роботи системи на рівні 87% та 92% відповідно.

Навіть враховуючи те, що двоетапний підхід дозволяє досягнути коректності роботи системи на рівні середнього кваліфікованого та досвідченого рентгенолога, він має свої недоліки. Перш за все, навчання згорткової нейронної мережі на етапі сегментації вимагає наявності відповідних міток в тренувальному наборі даних. Дані мають містити інформацію про область та регіон, де знаходиться підозрілі ущільнення і чи є вони злоякісними. В медицині, анотація і збір даних є завжди дуже дорогим та трудомістким процесом, що вимагає багато часу та грошей, тому анотація тренувальної вибірки є великою проблемою при розробці систем автоматичною діагностики раку.

По друге, етап класифікації, на якому вирішується чи містить область злоякісне утворення, втрачає глобальну інформацію про легені, положення даної області відносно всього органу. Така інформація може допомогти покращити розуміння того, чи містить конкретна область злоякісне утворення чи ні. Більше того, двоетапна система не дозволяє аналізувати та додавати історичну інформацію, наприклад дані із попереднього знімку пацієнта. Такого роду інформація дозволяє оцінити, чи присутні утворення, що з часом збільшилися у розмірі. Якщо такі утворення існують, то це говорить про наявність в цьому місці злоякісної пухлини з високою долею ймовірності.

Описані вище проблеми можуть бути вирішені, якщо пайплайн системи діагностики розбити на ще більше етапів. Саме такий сценарій був реалізований в роботі [94], що описує систему автоматичної діагностики раку легенів,

яка складається із чотирьох основних етапів, як показано на рис. 2.7.

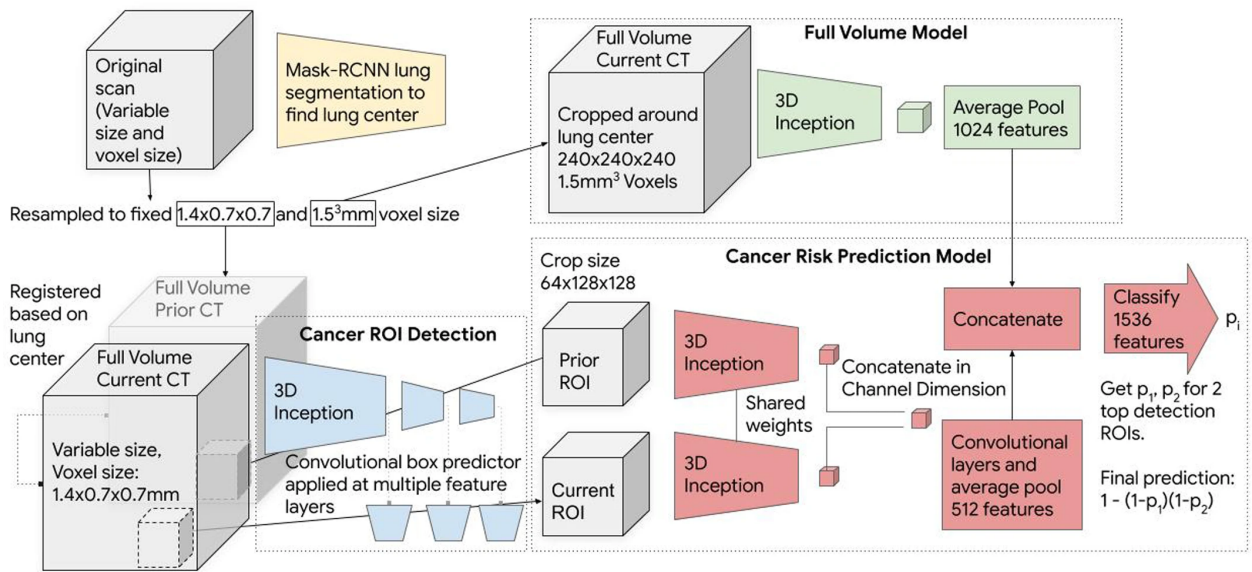


Рис. 2.7. Багатоетапна система діагностики раку легенів. Зображення взято з [94].

На першому етапі системи виконується сегментація вхідного тривимірного зображення легенів пацієнта. Сегментація виконується за допомогою тривимірної згорткової нейронної мережі [95]. Основна мета цього етапу полягає у відділенні на зображенні легеневої тканини від інших тканин, таких як кістки чи судини. Основна відмінність від попередніх систем полягає в тому, що тут не виконується попередня обробка КТ зображення окрім як зміни його розміру. На другому етапі використовується інша тривимірна згорткова нейронна мережа, яка знаходить області із потенційно злоякісними утвореннями. Ця процедура реалізується подібного до того, як це робиться на першому етапі в описаній вище двоетапній системі. На цьому ж етапі на вхід нейронної мережі може подаватися попередній знімок пацієнта для знаходження потенційно злоякісних областей на старому знімку. На останньому етапі ця інформація буде використовуватися для винесення висновку про наявність раку легенів у пацієнта. Третій етап використо-

вує тривимірну згорткову нейронну мережу, яка намагається класифікувати наявність раку легенів напряду, використовуючи зображення легенів отримане після першого етапу. Тривимірна згорткова нейронна мережа використовується і навчається схожим чином, як описувалося в одноетапному підході. Тут відмінність полягає у тому, що коли навчання закінчуються, всі не згорткові шари відкидаються та замінюються на шар субдискретизації за середнім значенням. Ознаки з цього шару згодом подаються на наступний фінальний етап. На фінальному етапі використовується тривимірна згорткова нейронна мережа, що на вхід отримує підозрілі області з другого етапу поєднані із ознаками отриманими на третьому етапі. Також, якщо наявний попередній знімок пацієнта, а також на другому етапі були згенеровані потенційно злоякісні регіони для старого знімку, то ці регіони також обробляються. Спочатку, базові згорткові шари дістають ознаки із минулого знімку паралельно до ідентичної обробки поточного знімку. В кінці, ознаки отримані з потенційно злоякісних регіонів попереднього та поточного знімку поєднуються з ознаками мережі третього етапу. Після об'єднання ознак додаються ще декілька шарів нейронної мережі таким чином, щоб на виході поверталася ймовірність присутності раку легенів у пацієнта.

Чотирьохетапний підхід зміг ще більше покращити коректність роботи системи, та обійшов показники точності роботи середньостатистичного рентгенолога. Така система показала коректність роботи системи на рівні 95%, що показує ефективність такої системи та значне покращення роботи після збільшення кількості етапів в середині системи. Один із недоліків такої системи полягає у тому, що вона є дуже великою та займає велику кількість пам'яті, через те що складається з декількох достатньо великих тривимірних згорткових нейронних мереж. Також, така кількість етапів збільшує час, який необхідний системі для обробки КТ знімку пацієнта.

2.3. Висновки до розділу

В рамках цього розділу було проаналізовано архітектуру тривимірних згорткових нейронних мереж у порівнянні з класичними двовимірними згортковими нейронними мережами. Було показано, що тривимірні згорткові нейронні мережі можуть використовувати додаткову просторову інформацію доступну в тривимірних даних для прийняття рішення, в той час як двовимірні згорткові нейронні мережі втрачають її. Однак, тривимірні згорткові нейронні мережі використовують значно більшу кількість пам'яті та вимагають значно більшу обчислювальну потужність для виконання необхідних розрахунків. Для навчання тривимірних нейронних мереж можна адаптувати всі існуючі підходи, які використовуються для навчання двовимірних згорткових нейронних мереж

Також, було показано, що для побудови системи автоматичної діагностики раку легенів найкраще використовувати багатоетапні системи, що складаються з набору окремих тривимірних згорткових нейронних мереж, що виконують свою власну підзадачу. Так, задача діагностики може бути розділена на етапи сегментації легенів та злоякісних утворень на знімку, класифікації отриманих регіонів злоякісних утворень. Такі системи можуть досягати коректності роботи системи на рівні 90% та значенням ROC AUC на рівні 95%.

Основні результати описані в даному розділі були опубліковані в наукових працях [5, 7, 96].

РОЗДІЛ 3

РЕКУРЕНТНІ НЕЙРОННІ МЕРЕЖІ ДЛЯ ДІАГНОСТУВАННЯ РАКУ ЛЕГЕНІВ

Обговорені в попередньому розділі тривимірні згорткові нейронні мережі дозволяють створити систему автоматичної діагностики раку легенів коректність роботи якої буде близьким до коректності роботи експерта. Однак, такі мережі мають високі вимоги до обчислювальної потужності та наявності великого об'єму пам'яті. Можливість використання альтернативного підходу, що використовує двовимірну операцію згортки але в той же час може розуміти просторові залежності між різними шарами КТ знімку може допомогти зменшити ці вимоги.

Для розуміння альтернативного підходу потрібно поглянути на те, як експерт аналізує КТ знімки легенів пацієнта. Зазвичай, професійний рентгенолог проглядає двовимірне зображення на кожному шарі КТ знімку, намагаючись віднайти потенційно злоякісну пухлину або невеликі підозрілі ущільнення. Перегляд та аналіз знімку зазвичай займає близько 10 хвилин у досвідченого рентгенолога. Кожен шар уважно і детально аналізується, порівнюється з попередніми та наступними шарами КТ знімку. Якщо знайдено потенційне підозріле ущільнення, інформація про нього запам'ятовується та сама локація перевіряється на попередніх та наступних шарах знімку. Інформація між цими зображеннями, що знаходяться на різних шарах КТ знімку, поєднується, вивчаються просторові залежності в підозрілих локаціях, аналізується розміщення, форма та морфологія такого ущільнення. Після цього, експерт робить висновок про те, чим являється дане ущільнення та чи є воно небезпечним для пацієнта.

Подібний підхід та механіку аналізу КТ знімку пацієнтів можна реалі-

зувати використовуючи штучні нейронні мережі. Для аналізу послідовності зображень кожного шару КТ знімку можна використовувати рекурентну нейронну мережу, що є класичною моделлю для роботи з послідовностями. Також, для виявлення більш складних патернів на кожному шарі індивідуально можна використовувати згорткову нейронну мережу. В цьому випадку нейронна мережа буде використовуватися як функція екстрактор високорівневих ознак для кожного окремого шару КТ знімку, послідовність яких буде використовуватися та аналізуватися рекурентною нейронною мережею. Рекурентна нейронна мережа буде повертати ймовірність присутності раку легенів на кожному із знімків.

Далі в розділі розглянемо та проаналізуємо детально рекурентні нейронні мережі, їх види, особливості роботи та навчання. Також, буде розглянуто способи та механізми, що допомагають нейронній мережі звертати увагу тільки на необхідні області КТ знімку.

3.1. Рекурентні нейронні мережі

Під рекурентними нейронними мережами (РНМ) розуміють клас нейронних мереж, основною метою при розробці яких є можливість обробки послідовностей. Особливість рекурентних нейронних мереж полягає в можливості обробляти довгі послідовності та можливість роботи з послідовними наборами даних різної довжини. На відмінну від традиційних нейронних мереж, рекурентні нейронні мережі мають властивість пам'яті та не розглядають кожне вхідне значення як щось нове та окреме. Така властивість може бути корисною в багатьох сценаріях. Наприклад, коли ми читаємо книжку нам необхідно пам'ятати та розуміти, що ми читали в минулому, для того щоб зрозуміти про що книжка розповідає на поточній сторінці. Більше того, в багатьох мовах світу деякі слова можуть мати різні значення в залежності від контексту, тому можливість зберігати та

використовувати інформацію із минулого досвіду є надзвичайно корисною і відкриває можливості для роботи з великим класом задач і проблем.

Можливість рекурентних нейронних мереж працювати з довгими послідовностями та запам'ятовувати історичну інформацію досягається шляхом додавання циклічних залежностей до нейронів. Тобто, один шар нейронної мережі має зворотній зв'язок на стан тієї самої нейронної мережі на попередньому кроці. Такі значення рекурентної мережі називають прихованим станом (англ. hidden state). Часто, роботу рекурентних нейронних мереж можливо описати наступним чином:

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; W), \quad (3.1)$$

де t означає поточний крок послідовності в момент часу t ; $h^{(t-1)}$ - стан нейронної мережі на попередньому кроці в момент часу $t - 1$; $h^{(t)}$ - позначає новий прихований стан нейронної мережі, що був розрахований для моменту часу t ; $x^{(t)}$ позначає зовнішній вхідний сигнал, що подається на вхід нейронної мережі в момент часу t ; W - вагові коефіцієнти нейронної мережі, що регулюють взаємозв'язок між прихованими станами нейронної мережі; функція f представляє функцію, яка відображає прихований стан нейронної мережі на момент часу t в наступний стан $h^{(t+1)}$.

З виразу 3.1 можна помітити, що рекурентна нейронна мережа може запам'ятовувати надану зовнішню інформацію x з попередніх кроків шляхом формування певного стану h . Коли така мережа навчається прогнозувати майбутні значення базуючись на отриманих минулих результатах, вона вчиться використовувати свій прихований стан в якості узагальнення всіх даних, що були отримані до поточного часу t . Такий прихований стан є високорівневим узагальненням історичної інформації та при збереженні його, модель може втрачати багато деталей, що були присутні в оригінальних даних. Залежно від критеріїв навчання, така нейронна мережа може навчитися звертати увагу на деякі аспекти значно більше, ніж на інші. Та-

ким чином буде зберігатися тільки та інформація, яка дійсно важлива для виконання поставленої задачі.

Цікаво, що рекурентна нейронна мережа може мати різну кількість прихованих станів в залежності від довжини послідовності, що подається на вхід. Часто, рекурентну нейронну мережу не зручно представляти рекурсивною функцією, тому що не зрозуміло наскільки глибокою є дана мережа, а сама структура є динамічною. В такому випадку використовують розгорнуте представлення рекурентної нейронної мережі. Якщо вхідна послідовність має скінченну кількість кроків T , то розгорнутий граф рекурентної нейронної мережі може бути побудований шляхом застосування визначення мережі, описаного у виразі 3.1, $T - 1$ раз. Графічно, процедуру розгортання графу рекурентної нейронної мережі показано на рисунку 3.1.

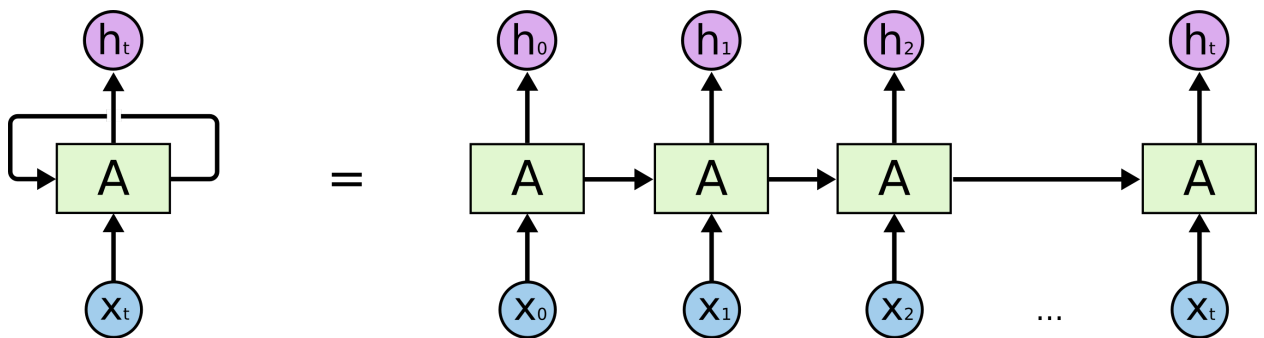


Рис. 3.1. Приклад відображення рекурентної нейронної мережі. Зліва зображено представлення у вигляді згорнутого графа, тоді як праворуч відображено відповідний розгорнутий граф. Зображення взято з [97].

Рекурентні нейронні мережі та їхня процедура розгортання графу має декілька основних переваг [47]:

1. Розмірність вхідних величин завжди фіксована та не залежить від довжини самої послідовності. Так відбувається через те, що ми задаємо структуру мережі в термінах переходу з одного стану в інший.

Якби ми використовували традиційні нейронні мережі, то ми б задавали структуру нейронної мережі в термінах довжини послідовності. В такому випадку нам би довелося навчати окрему нейронну мережу для кожної довжини послідовності, а потім об'єднувати результати роботи всіх таких нейронних мереж.

2. Завдяки структурі рекурентної нейронної мережі є можливість використовувати одну й ту саму функцію переходу f з одними й тими ж параметрами та ваговими коефіцієнтами на кожному з кроків послідовності. Така особливість допомагає нейронній мережі краще узагальнюватися до послідовностей відмінної довжини, які відсутні в тренувальній вибірці. Також, завдяки цьому рекурентна нейронна мережа вимагає менше прикладів в тренувальній вибірці у порівнянні з мережами, що не мають спільних вагових коефіцієнтів.

Рекурентна нейронна мережа може без проблем використовуватися для апроксимації вхідних даних. Це підтверджується теорією універсальної апроксимації, яка також була доведена для рекурентних нейронних мереж. Згідно теорії універсальної апроксимації, рекурентна нейронна мережа з достатньою кількістю прихованих станів може апроксимувати будь-яке вимірювальне відображення послідовності в іншу послідовність з необхідною коректністю (англ. accuracy) [98].

Навчання рекурентної нейронної мережі складається із двох основних етапів, як і в традиційних нейронних мережах: етап прямого розповсюдження (англ. forward propagation) та етапу зворотнього розповсюдження (англ. backpropagation), в рамках якого розраховуються градієнти для градієнтного спуску. В загальному випадку, процес прямого розповсюдження для рекурентної нейронної мережі можна записати наступними рівняннями:

$$h^{(t)} = \sigma(b + Wh^{(t-1)} + Ux^{(t)}), \quad (3.2)$$

$$o^{(t)} = c + Vh^{(t)}, \quad (3.3)$$

$$\hat{y}^{(t)} = g(o^{(t)}), \quad (3.4)$$

де b та c - вектори зміщення; U - матриця вагових коефіцієнтів для поєднання вхідних значень та приховано стану; V - матриця вагових коефіцієнтів для поєднання значень приховано стану та виходу нейронної мережі; W - матриця вагових коефіцієнтів для поєднання вхідних значень та приховано стану; σ - функція активації, яка зазвичай в рекурентних нейронних мережах має вигляд гіперболічного тангенсу; g - функція активації, що застосовується на виході із рекурентної нейронної мережі. Вигляд цієї функції відрізняється від типу задачі та нейронної мережі, що використовується. Якщо вирішується задача при якій повинна повертатися послідовність, то в такому випадку як функцію активації можна розглядати функцію softmax.

Після розрахованого значення $o^{(t)}$ можна розраховувати функцію втрат для рекурентної нейронної мережі, що буде мінімізуватися для забезпечення процесу навчання мережі. Наприклад [47], для рекурентної нейронної мережі, що будує вихідну послідовність для певної вхідної послідовності, функція втрат може мати наступний вигляд:

$$\begin{aligned} L(\{x^{(1)}, \dots, x^{(T)}\}, \{y^{(1)}, \dots, y^{(T)}\}) &= \sum_t L^{(t)} = \\ &= - \sum_t \log p_{model} \left(y^{(t)} | \{x^{(1)}, \dots, x^{(t)}\} \right), \end{aligned} \quad (3.5)$$

де $p_{model} \left(y^{(t)} | \{x^{(1)}, \dots, x^{(t)}\} \right)$ розраховується шляхом читання кроку t послідовності y з виходу рекурентної нейронної мережі $\hat{y}^{(t)}$.

Градiєнтний спуск розраховується по функції втрат, що описана вище. Градiєнти будуть розраховуватися стандартним алгоритмом зворотнього розповсюдження помилки (англ. backpropagation) по розгорнутому графу рекурентної нейронної мережі. Через те, що в цьому випадку алгоритм зворотнього розповсюдження виконується по всім часовим крокам вхідної

послідовності, його називають алгоритмом зворотнього розповсюдження по часу (англ. backpropagation through time).

Використовуючи процес розгортання графу є можливість будувати різноманітні архітектури рекурентних мереж в залежності від типу задачі, яку необхідно вирішити. Виділяють декілька основних видів рекурентних нейронних мереж [47]:

- Рекурентні нейронні мережі, що мають рекурсивний зв'язок між прихованими станами та повертають значення на кожному із кроків вхідної послідовності.
- Рекурентні нейронні мережі, що мають рекурсивний зв'язок між виходом нейронної мережі з попереднього кроку та поточним прихованим станом. Така нейронна мережа повертає відповідь на кожному кроці вхідної послідовності.
- Рекурентна нейронна мережа, що має рекурентні зв'язки між прихованими станами та на всю вхідну послідовність повертає тільки одне значення. Значення повертається тільки після того як вся вхідна послідовність була оброблена.

Якщо взяти до уваги проблему діагностики раку легенів, яка вирішується в даній дисертації, то для вирішення цієї задачі необхідно орієнтуватися на третій тип рекурентних нейронних мереж. Причина цьому полягає у тому, що вхідні дані містять послідовність зображень із різних шарів КТ знімку грудної клітини пацієнта, в той час як вихідний результат роботи системи повинен повертати тільки одну фінальну відповідь — ймовірність наявності раку легенів на знімку. Однак, описані вище рекурентні моделі не підходять для аналізу шарів КТ знімку пацієнта, оскільки вони вміють дивитися тільки на історію в минулому. При аналізі кожного шару зображення необхідно мати інформацію не тільки про ті злоякісні утворення, що були знайдені на минулих кроках, але й інформацію про утворення на наступних шарах. Якщо рекурентна нейронна мережа зможе одночасно

слідкувати за зображеннями з минулих та майбутніх кроків послідовності, буде можливо краще оцінити морфологію таких утворень та контекст. Така додаткова інформація допоможе підвищити коректність роботи моделі.

3.2. Двонаправлені рекурентні нейронні мережі

Для вирішення подібних проблем, були розроблені так звані двонаправлені рекурентні нейронні мережі [99] (англ. *bidirectional recurrent neural networks*). Основна ідея такої нейронної мережі полягає в формуванні навчальної послідовності з двох частин, що використовуватиметься на різних прихованих рекурентних шарах. Навчальна послідовність складається з послідовності прямого розповсюдження, де зберігається порядок елементів від початку до кінця, та з оберненої послідовності, яка відтворює ту саму вхідну послідовність з кінця до початку. Такі дві послідовності сприймаються двома різними рекурентними шарами, виходи яких поєднуються на вихідному шарі. Завдяки такій структурі вихідний шар має інформацію як і про вже пройдені кроки, так і про майбутні кроки кожної точки заданої послідовності, яка використовується на етапі тренування нейронної мережі. Маючи всю необхідну інформацію, вихідний шар може використовувати тільки найбільш релевантні ознаки з минулих та майбутніх кроків [100]. Схематичне зображення розгорнутого графу двонаправленої рекурентної мережі представлено на рис. 3.2. Алгоритм прямого розповсюдження двонаправленої рекурентної нейронної мережі описаний в алгоритмі 1.

Алгоритм зворотнього розповсюдження помилки для двонаправленої рекурентної нейронної мережі залишається незмінним. Єдина різниця полягає в тому, що градієнти спочатку розраховуються для вихідного шару, а потім вже для двох прихованих шарів, що відповідають за прямий та обернений порядок послідовностей.

Двонаправлені рекурентні нейронні мережі добре себе зарекомендували

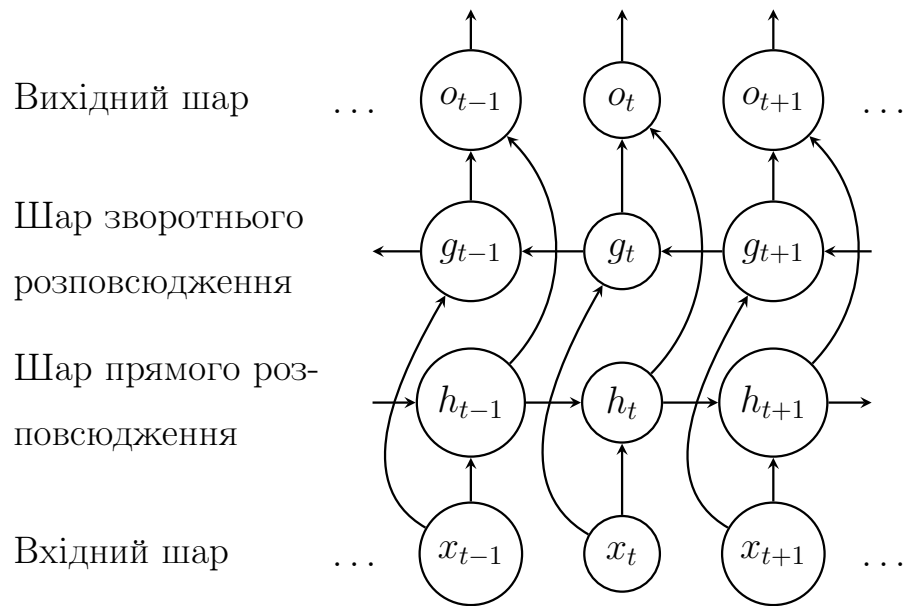


Рис. 3.2. Приклад розгорнутого графу двонаправленої рекурентної нейронної мережі.

for $t = 1$ to T **do**

 Проходження послідовності в заданому напрямі в шарі прямого розповсюдження на кожному кроці.

end for

for $t = 1$ to T **do**

 Проходження послідовності в оберненому напрямку від кінця до початку в зворотньому прихованому шарі нейронної мережі.

end for

for all t , in any order **do**

 Обробка відповідей з двох попередніх етапів на вихідному шарі. Тут враховується прихований стан з обох шарів.

end for

Алгоритм 1. Алгоритм прямого розповсюдження для двонаправленої рекурентної нейронної мережі [100].

в різних доменах, наприклад в задачах розпізнавання мови [101] та біоінформатики [102]. З того часу було опубліковано велику кількість робіт,

які показували що двонаправлені рекурентні нейронні мережі працюють значно краще за односторонні.

Однак звичайні рекурентні нейронні мережі мають один фатальний недолік, який не дозволяв отримати від них ще більш кращих результатів і чому прості рекурентні нейронні мережі не будуть працювати на КТ знімку. Проблема таких рекурентних нейронних мереж пов'язана з довгостроковими та короткостроковими залежностями. Часто, при аналізі послідовностей необхідно розуміти контекст, інформація про який може знаходитися на більш ранньому кроці. Як показує практика звичайні рекурентні нейронні мережі погано працюють з такими довгими залежностями і мають тенденцію забувати те, що було раніше, але краще пам'ятають те, що було в найближчих кроках послідовності. Проблема стандартних рекурентних нейронних мереж полягає в тому, що чим більша послідовність, тим більша кількість рекурентних зв'язків, що призводить до проблеми градієнтного вибуху або затухання [100] при навчанні градієнтними методами. Причина виникнення цієї проблеми полягає в тому, що при розгортанні рекурентної нейронної мережі для дуже довгих послідовностей ми отримаємо достатньо глибоку нейронну мережу. Як ми розглядали в попередніх розділах, глибокі нейронні мережі не можуть навчатися без нормалізації проміжних шарів, через те що з проходженням сигналу по структурі нейронної мережі градієнти значення або значно зростають, або значно зменшуються. Для вирішення цієї проблеми в стандартних рекурентних нейронних мережах був розроблений спеціальний тип рекурентних нейронних мереж з вентильним вузлом (англ. gated recurrent neural network).

3.3. Рекурентні нейронні мережі з вентильним вузлом

Найбільш популярною архітектурою рекурентної нейронної мережі з вентильним вузлом є LSTM [103](Long Short-Term Memory). Архітектура

цієї мережі складається з блоків із внутрішньою пам'яттю, які рекурентно поєднані між собою. Такі зв'язки є подібними до того, що ми маємо в стандартних рекурентних нейронних мережах, і інколи ще можуть називатися зовнішніми зв'язками LSTM. Кожен блок містить в собі інформацію про стан $s_i^{(t)}$, що має внутрішні рекурентні зв'язки до самого себе. Також, блок містить три основних контролери, які дають можливість зчитувати збережений стан з пам'яті, записувати або затирати інформацію [100]. Приклад архітектури LSTM зображений на рис 3.3.

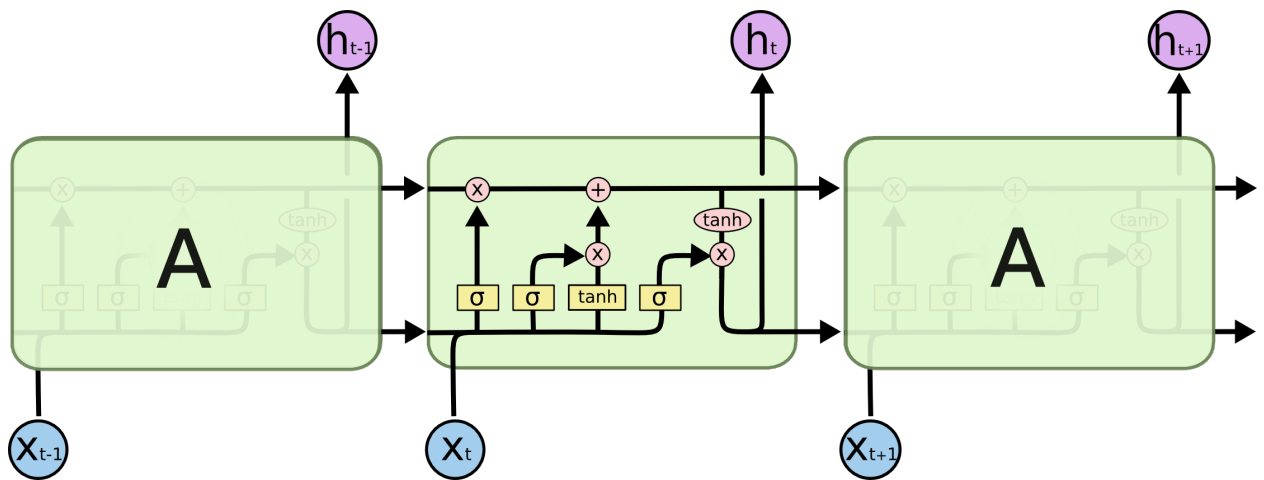


Рис. 3.3. Проста діаграма, що відображає принцип роботи прихованого шару LSTM. Зображення взято з [97].

Прямий прохід через один блок LSTM складається із декількох основних етапів, які по своїй суті виконують взаємодію із збереженим внутрішнім станом. Перший етап складається з так званого вентиля забуття (англ. forget gate unit), який вирішує яку інформацію необхідно затерти у внутрішньому стані, що зберігає інформацію отриману зі всіх попередніх кроків послідовності. Процес “забування” інформації відбувається шляхом використання сигмоїдної функції та матриці вагових коефіцієнтів [47]:

$$f_i^{(t)} = \sigma \left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)} \right), \quad (3.6)$$

де $x^{(t)}$ є поточним вхідним вектором на кроці t послідовності; $h^{(t)}$ - поточний вектор прихованого шару станом на поточний крок послідовності та містить вихідну інформацію LSTM блоків з попередніх кроків; b_i^f - зміщення вентиля забуття (англ. forget gate); U^f - вхідні вагові коефіцієнти вентиля забуття (англ. forget gate); W^f - рекурентні вагові коефіцієнти вентиля забуття (англ. forget gate).

Наступний етап блоку LSTM складається із декількох проміжних етапів. Спочатку визначається інформація про те, які значення в збереженому стані мають бути оновлені за допомогою так званого вхідного вентиля (англ. input gate). Далі, формується список нових елементів, що відображають нову інформацію яка повинна бути додана до вже існуючого стану. Останнім кроком поєднується отриманні значення на всіх етапах та оновлюється внутрішній стан $s_i^{(t)}$. Всі ці операції можуть бути описані наступними математичними виразами [47]:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma \left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)} \right), \quad (3.7)$$

$$g_i^{(t)} = \sigma \left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)} \right), \quad (3.8)$$

де b - вектор зміщення для всього LSTM блоку; U - вагові коефіцієнти для входу LSTM блоку; W - вагові коефіцієнти для рекурентних зв'язків між різними LSTM блоками на різних кроках послідовностей; $g_i^{(t)}$ - функція зовнішнього вхідного вентиля (англ. external input gate).

Останній етап в LSTM блоці вирішує, яку інформацію потрібно повертати. Вихідне значення розраховують за допомогою вихідного вентиля (output gate), що також використовує сигмоїдну функцію, але з власними ваговими коефіцієнтами:

$$h_i^{(t)} = \tanh \left(s_i^{(t)} q_i^{(t)} \right), \quad (3.9)$$

$$q_i^{(t)} = \sigma \left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)} \right), \quad (3.10)$$

де b^o , U^o , W^o - зміщення, вхідні та рекурентні вагові коефіцієнти вихідного вентилля відповідно.

Архітектура LSTM добре зарекомендувала себе на реальних задачах і показує, що мережа значно краще запам'ятовує та обробляє довгострокові залежності за стандартні рекурентні нейронні мережі [101]. Більше того, LSTM може бути двонаправленою, так само як і стандартні архітектури. Для цього потрібно зробити все те саме, що і описувалося в розділі 3.2, а саме використати два різні приховані LSTM блоки, один з яких працюватиме з прямою послідовністю, а інший з елементами послідовності сформованих в протилежному напрямку. Вихідне значення для такої нейронної мережі може вираховуватися так само, як і для стандартних двонаправлених рекурентних нейронних мереж.

Враховуючи успіх архітектури LSTM, а також можливість використання режиму двонаправленої рекурентної нейронної мережі роблять цю архітектуру одним із найкращих варіантів для вирішення задачі автоматичної діагностики раку легенів. Двонаправлену LSTM можна модифікувати для роботи із двовимірними зображеннями шляхом роботи з послідовностями пікселів з ліва на право, з право наліво, знизу доверху та з верху донизу, призначивши на кожну з них окремий рекурентний LSTM блок. Однак, такий підхід буде вимагати великі обчислювальні затрати, та ускладнює збіжність нейронної мережі. Тож можна використати згорткову нейронну мережу для роботи з двовимірними вхідними даними, а замість повноз'язного шару можна використати двонаправлену LSTM. Однак, така нейронна мережа однаково буде складною для навчання, особливо враховуючи, що вибірка яка використовується в даній роботі не містить багато екземплярів на тренувальному наборі даних. Для покращення та пришвидшення навчання нейронної мережі можна скористатися так званим

механізмом уваги, який може допомогти нейронній мережі локалізувати місцезнаходження злоякісних утворень, що потенційно можуть вказувати на наявність раку легенів на знімку.

3.4. Механізм уваги в задачах комп'ютерного зору

Механізм візуальної уваги є одним із основних інструментів багатьох аспектів складного процесу мислення людини, у тому числі в процесі осмислення візуальної інформації [104]. Наприклад, при аналізі та перегляді КТ знімку легенів пацієнту рентгенолог не дивитися одночасно на всі локації зображення. Замість цього, людина фокусує свою увагу на певних локаціях рентгенівського знімку. Спочатку, людина проходить по всьому знімку частинами, по різних локаціям, знаходячи найбільш підозрілі легеневі утворення. Далі, знайдені локації аналізуються більш детально шляхом фокусування зору на конкретному місці. Фокусуючись на певній локації, людина концентрує свою увагу на більш дрібних деталях, можливо час від часу переводячи погляд на сусідні структури. Таким чином, людина більш обчислювано ефективно працює з візуальною інформацією на відмінну від штучних нейронних мереж, а її зорове сприйняття не залежить від розміру та масштабу вхідної перспективи.

Глибинні нейронні мережі дивляться та обробляють візуальні дані в цілому, не розбиваючи їх на менші підмножини. Таким чином, при побудові моделі для роботи із зображеннями завжди будуть серйозні обмеження, оскільки чим більше вхідне зображення, тим більше пам'яті та обчислювальної потужності необхідно для його обробки. Однак, будь-яка глибинна нейронна мережа має свій неявний внутрішній механізм уваги (англ. *implicit attention*), який концентрує увагу моделі на певних локаціях вхідного зображення. Рисунок 2.4 показує, що глибинна згорткова нейронна мережа на кожному шарі дивиться на конкретні патерни та виділяє відповідні

ознаки. На останніх шарах мережі можливо навіть побачити патерни, які дуже нагадують об'єкт, який така мережа може класифікувати.

Таким чином, згорткова нейронна мережа виділяє певні конкретні області на зображенні, тоді як інші області, що не відносяться до об'єкта розпізнавання, просто ігноруються або мають дуже малий вплив. Наприклад, на рисунку 3.4 зображено, як виконуючи задачу опису зображення ней-

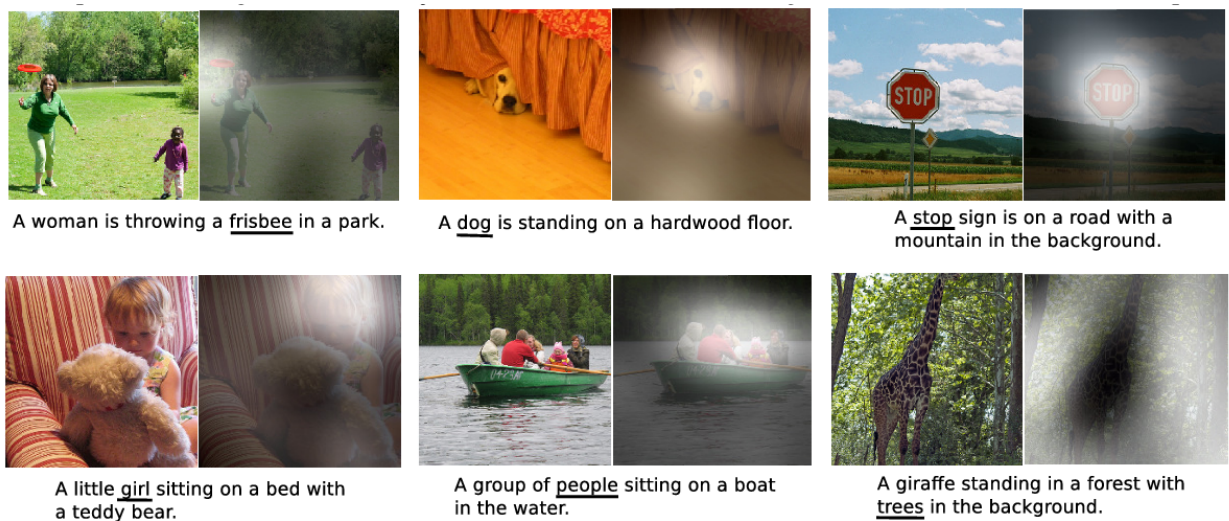


Рис. 3.4. Приклад механізму уваги в нейронних мережах, що застосовуються для вирішення задачі створення опису зображення [105].

ронна мережа концентрується на конкретному об'єкті, що знаходиться в конкретній області та відповідає текстовому слову. Схожа ситуація може спостерігатися і для задач класифікації об'єкта, що знаходиться на зображенні. Таку властивість глибинних нейронних мереж можна назвати певною формою механізму уваги, якому вона навчилася на тренувальній вибірці.

Таке явище неявної уваги може пояснюватися тим, що глибинна нейронна мережа навчається розпізнавати об'єкти на вхідному зображенні шляхом збільшення або зменшення вагових коефіцієнтів. Іншими словами, нейронна мережа навчається більш сильно реагувати на певні локальні патерни на зображенні, чим на інші, тим самим виділяючи найбільш релевантні пікселі на зображенні. Таку властивість глибинних нейронних мереж часто вико-

ристовуюють для аналізу та інтерпретації їх роботи. Для відображення чутливих областей можна використовувати матрицю Якобі, що представляє собою інформацію про чутливість виходів моделі до того, що було подано на вхід:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_k} \end{bmatrix}, \quad (3.11)$$

де J - матриця Якобі, що розраховується під час виконання алгоритму зворотнього розповсюдження помилки; x - вхідний вектор розміром k ; y - вихідний вектор розміром m .

Таким чином нейронні мережі вже мають вбудований неявний механізм уваги, який може виділяти певні конкретні області на вхідних даних. Однак, людина може керувати своєю увагою, наприклад цілеспрямовано переводячи погляд на іншу локацію, або, наприклад, вичитавши якусь інформацію в довіднику, звернути увагу на певну особливість на зображенні. Подібні механізми уваги також намагаються реалізувати в нейронних мережах. В термінології глибокого навчання, такий механізм називається явним механізмом уваги (англ. explicit attention mechanism), через те що в такому випадку нейронні мережі явно та цілеспрямовано отримують додатковий сигнал про локацію на яку потрібно звернути увагу. Одним із прикладів застосування механізму уваги є робота [106], де намагалися класифікувати рукописні цифри на зображеннях використовуючи рекурентну нейронну мережу та працюючи на кожному кроці тільки з обмеженою частиною зображення.

До основних переваг використання явно заданого механізму уваги в нейронних мережах можна віднести [107]:

- Обчислювальну ефективність навчання та роботи такої системи. Завдяки механізму явної уваги можна вказати нейронній мережі на пріоритетні ознаки та області для виконувальної задачі, тим самим

пришвидшивши та покращивши коректність роботи моделі.

- Масштабованість. За рахунок механізму уваги можна не працювати відразу зі всім зображенням, а тільки виділити найбільш потенційно важливу область та працювати з нею.
- Послідовна обробка статичних даних.
- Краща інтерпретованість виходів моделі.

Ті чи інші переваги механізму явної уваги можуть сильніше або слабкіше проявлятися в залежності від типу механізму уваги, що використовуються для побудови моделі [108]. В основному розділяють два основних види механізму уваги в нейронних мережах: механізм м'якої (англ. *soft attention*) та жорстокої уваги (англ. *hard attention*). Найбільш розповсюдженим механізмом уваги є механізм м'якої уваги [109], оскільки такі моделі є повністю диференційованими та без проблем можуть навчатися за допомогою градієнтного спуску та алгоритму оберненого розповсюдження.

Нейронні мережі з механізмом м'якої уваги реалізується у вигляді адаптивної матриці вагових коефіцієнтів (маски) по певним шарам згорткової нейронної мережі. Такий підхід допомагає покращити коректність роботи мережі шляхом виділення та ізолювання важливої інформації. Неважлива інформація не обробляється на наступних шарах нейронної мережі, хоч вона й присутня в даних. Не враховується така інформація шляхом прирівнювання вагових коефіцієнтів до нуля алгоритмом машинного навчання. Навчання нейронної мережі стає більш ефективним та вимагає меншої кількості даних в тренувальній вибірці, тому що складність взаємодії між різними частинами інформації спрощується за рахунок відсутності неважливих елементів. Додатковими перевагами підходу також є простота концепції та повна диференційованість моделі, що дозволяє без проблем використовувати градієнтні методи навчання. Недоліком м'якого механізму уваги є те, що на практиці не всі вагові коефіцієнти неважливих даних будуть дорівнювати нулю. Це призводить до доступності нерелевантних

ознак разом із важливими, що може негативно вплинути на результат навчання мережі.

На противагу, жорсткий механізм уваги використовує тільки певну підмножину інформації, яку вважає найбільш релевантною. Всі нерелевантні ознаки відсіюються та не використовуються далі. Як і механізм м'якої уваги, механізм жорсткої уваги покращує ефективність навчання нейронної мережі шляхом виділення найбільш важливих значень у вхідних даних. Однак, механізм жорсткої уваги є більш ефективнішим за рахунок того, що неважлива інформація повністю виключається із розрахунків. Проте, серйозним недоліком такого підходу є те, що процес вибору найбільш релевантної інформації є дискретною операцією. В такому випадку ми маємо недиференційовану частину моделі, що означає градієнтний спуск не зможе оптимізувати механізм вибору інформації під час навчання. Для того, щоб обійти цю проблему можна спробувати звести задачу оптимізації механізму вибору релевантної інформації до задачі навчання із підкріпленням [106]. Або, деякі автори намагаються використовувати певну функцію припущення, що вибирає найбільш релевантні значення. Наприклад, такою функцією може бути l^2 норма [108]. Однак, способи обходу цієї проблеми до цих пір є активно сферою дослідження.

Враховуючи ефективність механізму уваги та те, що такий підхід може допомогти використати попередньо анотовані локації пухлини в КТ зображеннях легень пацієнтів, в рамках дисертації такий механізм було вирішено використати для покращення ефективності навчання власної рекурентної моделі.

3.5. Застосування рекурентної згорткової нейронної мережі з механізмом уваги для діагностики раку легенів

В попередніх розділах детально обговорювалася та описувалася інформація про рекурентні нейронні мережі та як вони допомагають працювати з послідовними даними. В цьому розділі описано більш детально модель, яку було побудовано для діагностування раку легенів на базі КТ зображень пацієнта. Опис даних та деталі навчання нейронної мережі описані в розділі 4.

Як вже обговорювалося в попередніх розділах, головною ціллю побудови рекурентної моделі є спроба досягти як можна більшої коректності роботи імітуючи підхід, який використовує лікар при аналізі КТ знімків легенів людини. Щоб досягти поставленої мети, для кожного зображення шарів КТ знімку вирішено використовувати двовимірну згорткову нейронну мережу та двонаправлену рекурентну нейронну мережу з вентильним вузлом, що має поєднувати ознаки отриманні на кожному з шарів. В якості двовимірної згорткової нейронної мережі була використана архітектура DenseNet [62], через те що така мережа дозволяє побудувати достатньо велику нейронну мережу використовуючи меншу кількість параметрів. Ще одним аргументом для використання цієї архітектури є те, що вона зарекомендувала себе краще за більш ранню архітектуру ResNet на задачах класифікації. Також, у вільному доступі наявні різні модифікації DenseNet, що навчені на наборі даних ImageNet. Попередньо навчена нейронна мережа необхідна для використання навчання перенесенням, що пришвидшить та спростить процедуру навчання на наявному наборі даних КТ зображень. В побудованій архітектурі згорткова нейронна мережа виступає таким собі екстрактором послідовностей високорівневих ознак з кожного зображення шару КТ знімку легень.

В якості двонаправленої рекурентної нейронної мережі з вентильним

вузлом була обрана архітектура LSTM [110], робота якої детально описувалася в розділі 3.3. Основна задача рекурентної нейронної мережі в цьому випадку є обробка послідовності високорівневих ознак, отриманих з останнього шару згорткової нейронної мережі. Ідея полягає в тому, що рекурентна нейронна мережа може віднайти корисні просторові ознаки у третьому вимірі та зможе досягти результатів схожих із тривимірними згортковими нейронними мережами.

Після реалізації та навчання описаної архітектури рекурентної нейронної мережі було отримано результати коректності роботи моделі на рівні 72%. Такі результати частково пересікаються з результатами одноетапних тривимірних згорткових мереж, але в цілому результат виявився значно гіршим. Проблема цієї архітектури полягала в тому, що для навчання такої мережі необхідно мати значно більшу кількість даних на тренувальній вибірці даних. Модель не могла за виділений час віднайти достатньо корисні ознаки для досягнення вищих результатів.

Слідуючи тій самій логіці, що і для тривимірних згорткових нейронних мереж, було вирішено спробувати або розбити задачу на менші частини, або спробувати надати рекурентній мережі додаткову інформацію про наявність пухлин на знімку, що були наявні в наборі даних. Для цього було вирішено скористатися механізмом уваги, описаним в розділі 3.4. Ціллю було використати механізм м'якої уваги, щоб отримати повністю диференційовану модель та мати можливість навчити модель від початку до кінця методом градієнтного спуску та алгоритмом зворотнього розповсюдження для знаходження градієнтів. В якості базової ідеї механізму уваги було використано модель, описану в роботі [111]. В рамках цієї роботи автори будували систему для розпізнавання пошкоджень в головному мозку людини на базі КТ знімків черепа людини. Основною особливістю побудованої нейронної мережі було використання механізму м'якої уваги [106], яка надавала інформацію нейронній мережі про локацію злоякісного утворення

на знімку під час етапу навчання. Локація злоякісного утворення бралася із набору даних, який попередньо був описаний професійними рентгенологами. Такий підхід значно покращував точність роботи нейронної мережі для задачі знаходження пошкоджень в головному мозку людини. Тому, було вирішено спробувати адаптувати цей підхід до задачі автоматичної діагностики раку легенів.

Механізм м'якої уваги реалізовувався шляхом перевикористання анотацій локацій знаходження пухлин в наявному наборі даних. Основна ідея полягає в тому, що наявна інформація має впливати на вагові коефіцієнти згорткової нейронної мережі на різних рівнях архітектури DenseNet. Подібно до [111], це реалізовано шляхом додачі блоків оберненої операції згортки, що відновлює зображення до розмірів маски та надає інформацію про те, в якій локації на думку нейронної мережі знаходиться пухлина. Такі блоки додаються після кожного DenseNet блоку вибраної архітектури, щоб контролювати розповсюдження виділених ознак протягом всієї моделі. Отримана архітектура зображена на рисунку 3.5.

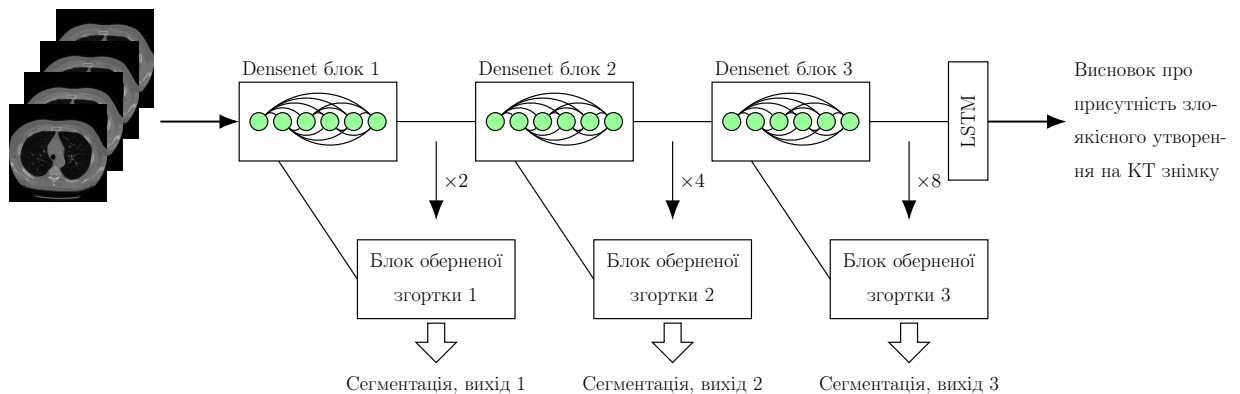


Рис. 3.5. Побудована рекурентна згорткова нейронна мережа з механізмом м'якої уваги. Для надання інформації про локацію пухлини використовується бінарна маска, що наявна в тренувальному наборі даних.

Інформація, яку повертає розроблена архітектура на кожному з блоків оберненої операції згортки використовується під час навчання моделі. Так,

отримана маска після кожного з блоків порівнюється з очікуваним результатом, який сформований на основі даних тренувальної вибірки. Протягом навчання мережі вагові коефіцієнти операції оберненої згортки та попередніх згорткових шарів в DenseNet блоці оновлюється таким чином, щоб отримувана маска як можна більше збігалася з маскою тренувального набору. Чим більш глибокі згорткові шари використовуються, тим більший вплив операція оберненої згортки має на згорткову нейронну мержу. Це досягається шляхом використання різних коефіцієнтів впливу до кожного виходу нейронної мережі, що задаються як гіперпараметри мережі перед її навчанням. На відмінно від [111], блоки оберненої операції згортки мають додатковий зв'язок з більш ранніми згортковими шарами, що мають схожу розмірність виходу. Такі зв'язки зроблені в такій ж манері, як в нейронній мережі типу U-Net [54] та нівелюють втрату інформації, яка відбувається за рахунок послідовного застосування операцій згортки та оберненої згортки. Також, для зменшення використання пам'яті моделі, вихідна маска має менший розмір за початкове вхідне зображення. Блок оберненої згортки складається з декількох послідовно з'єднаних шарів згортки 1×1 та оберненої згортки для побудови піксельної маски відображення локації пухлини.

Результатом використання механізму м'якої уваги стало значно вищий рівень коректності роботи моделі, що був отриманий на рівні 81%. Такий результат говорить про ефективність механізму уваги в побудові більш складних патернів для діагностування раку легенів. Цілком можливо, якщо покращити та ускладнити механізм уваги можна було б досягти ще кращих результатів.

3.6. Висновки до розділу

В рамках цього розділу було описано побудовану рекурентну нейронну мережу для діагностування раку легенів на КТ зображенні. Проведене дослідження показало, що рекурентні нейронні мережі працюють достатньо добре для того, щоб використовувати їх для розробки таких систем. Однак, навчання рекурентних мереж вимагають використання механізму уваги для збільшення ефективності навчання, без якого така модель покаже результат не вище за одноетапні тривимірні згорткові нейронні мережі. Однак, побудована рекурентна згорткова нейронна мережа не змогла досягти рівня коректності роботи системи на рівні багатоетапних тривимірних згорткових нейронних мереж. Однією з причин цього є менша ефективність навчання такої мережі, тобто така модель вимагає більшої кількості даних на тренувальній вибірці. Основним недоліком побудованої системи є вимоги до наявності інформації про локацію пухлини на знімку.

Основні результати отримані та описані в даному підрозділі були опубліковані в роботі [6].

РОЗДІЛ 4

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ДЛЯ ОБРОБКИ КТ ЗОБРАЖЕНЬ ЛЕГЕНІВ

4.1. Діагностика раку легенів. Опис даних

Рак легень – це захворювання, при якому злоякісна пухлина утворюється із клітин, які вистеляють внутрішню поверхню легень та бронхів [112]. В залежності від того, де розташована пухлина виділяють різні види раку легенів. Якщо пухлина зароджується в слизовій оболонці великих бронхів [4], то такий рак називають центральним. Рак легенів цього виду проявляється вже в самому початку розвитку пухлини, тому в більшості випадків його рано діагностують і своєчасно починають лікування. Другий вид раку легенів називають периферичним, оскільки він локалізується в дрібних бронхах та легеневої тканині. Основна небезпека другої форми полягає в тому, що така форма захворювання може протікати без серйозних симптомів [4]. Тому, в таких випадках рекомендують проводити регулярні обстеження, особливо для людей які знаходяться в зоні ризику (напр. курці, або ті, хто працює на шкідливому виробництві). Для діагностики раку легенів використовують флюорографію, рентгенографію, рентгеновську комп'ютерну томографію (КТ) та магнітно-резонансну томографію (МРТ), бронхоскопію [113]. В рамках дисертаційної роботи використовуються рентгеновські КТ знімки для діагностування та перевірки присутності раку легенів, через те що даний метод дає більш точну інформацію про структури в легенях пацієнтів, на відмінно від флюорографії та широкоформатної рентгенографії.

Під комп'ютерною томографією (КТ) розуміють метод рентгенівського сканування, при якому рентгенівський промінь пошарово та поступово проходить через тканини людського тіла в різних напрямках, з різних кутів та положень. В результаті отримуються пошарове рентгенівське зображення необхідної частини людського тіла. Отримане зображення може переглядатися як і в 2D форматі, так і 3D форматі, в залежності від потреб лікаря-рентгенолога. В рамках дослідження раку легенів, КТ зображення робиться на рівні грудної клітки пацієнта та налаштовується на розпізнавання легеневої тканини. В результаті, отримується зображення, що містить інформацію про легеневу тканину пацієнта. Також, зображення містить інформацію про сусідні тканини, такі як кістки, судини, тощо (див. рис. 4.1).

В рамках дисертаційної роботи для навчання нейронних мереж використовувалися набори даних LUNA [115] та Data Science Bowl 2017 (DSB 2017) [114]. DSB 2017 містить результати КТ для більш ніж 1000 пацієнтів. Кожне зображення супроводжується інформацією про присутність раку легенів на знімку. Діагнози раку легенів в наборі даних були підтвердженні в лікарнях. DSB 2017 розбитий на тренувальну, перевірочну та тестову вибірки, що містять 1397, 198 та 506 зображень грудної клітки пацієнтів відповідно. Тренувальна вибірка є не збалансованою, та містить 1035 прикладів на яких відсутній рак, та 362 зображень-прикладів, що містить пухлину.

Набір даних LUNA [115] містить інформацію про результати КТ для 888 пацієнтів. Особливостями даного набору даних є те, що присутні анотації 1186 міток з координатами, що позначають область з потенційно злоякісною пухлиною. Дані були створені і описані професійними рентгенологами.

Для діагностування раку легенів, рентгенологи шукають області, в яких знаходяться злоякісні утворення, які утворюють потенційні пухлини. Зазвичай, для діагностування раку намагаються віднайти всі видимі на знімку утворення, навіть якщо вони є не дуже великими. Небезпечними і най-

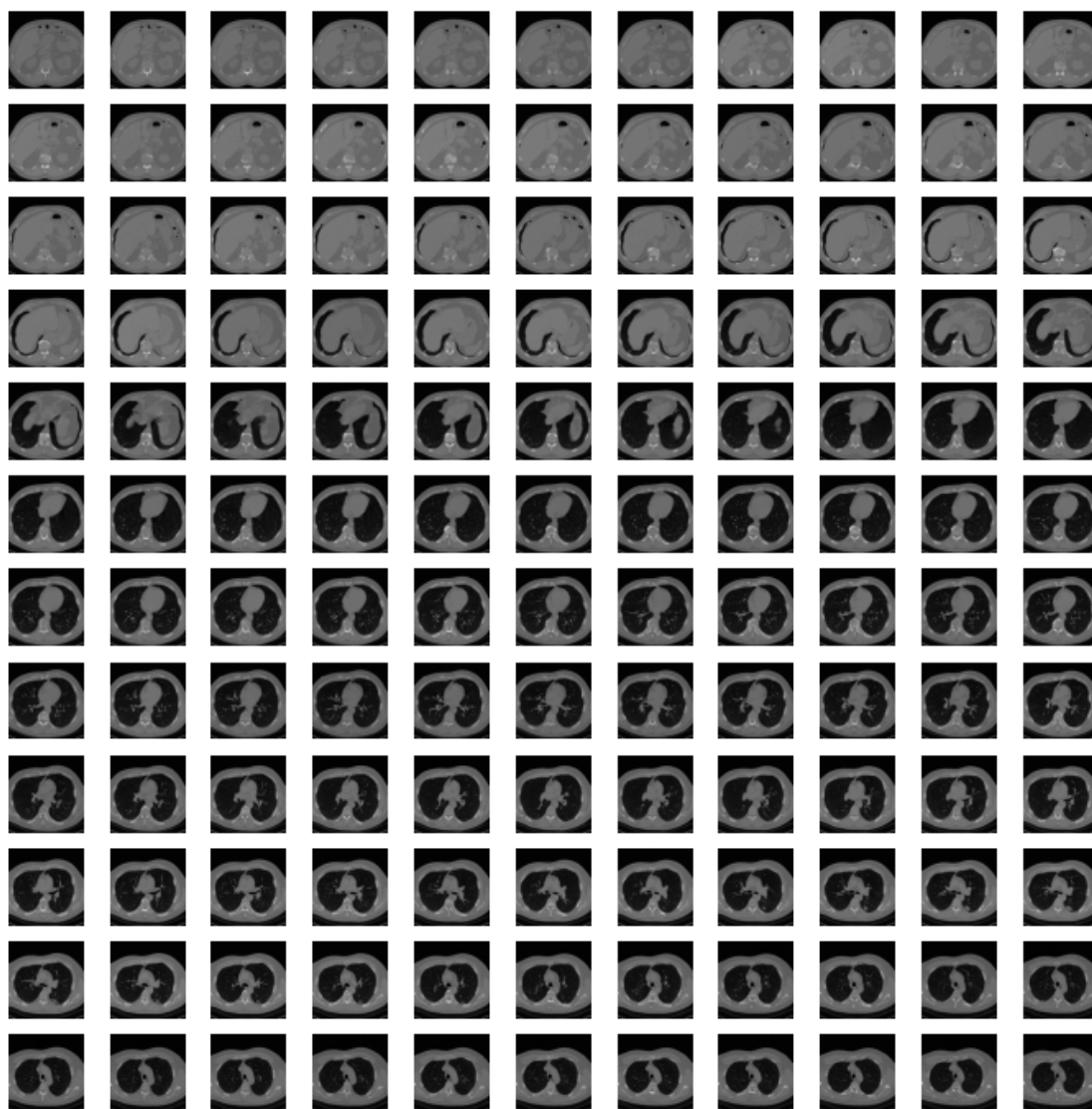


Рис. 4.1. Приклад КТ знімку легенів пацієнта, взятий із набору даних DSB 2017 [114].

більш ймовірними злоякісними пухлинами вважаються утворення розміром 6 мм [116] і більше. Якщо ж утворення невелике, можуть призначити додаткові аналізи, або попросити пацієнта пройти повторну діагностику протягом півроку або року. При наступній діагностиці буде перевірятися чи знайдені утворення не збільшилися протягом часу. Якщо такий вузлик значно збільшився у розмірах, говорять про велику ймовірність наявності раку легенів та призначають біопсію для більш точного діагностування. Тому, при побудові системи автоматичної діагностики раку легенів, часто

додають етап сегментації злоякісних утворень для надання лікарям більше інформації про причини діагностування чи не діагностування раку.

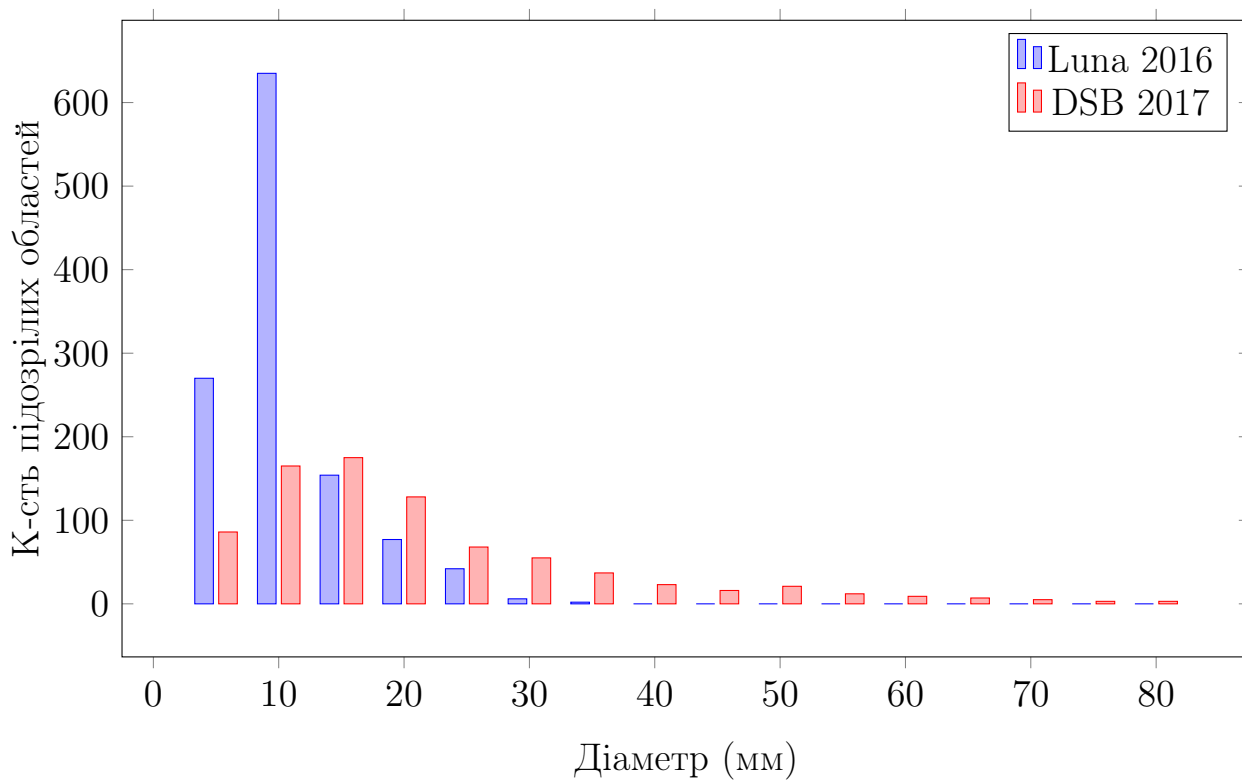


Рис. 4.2. Порівняння розподілу діаметру потенційних пухлин в LUNA та DSB 2017 наборах даних.

Для виконання задачі сегментації регіонів із злоякісними утвореннями набір даних має використовувати попередньо анотовані КТ зображення, де зазначаються координати такого регіону. Набір LUNA містить таку інформацію, проте набір даних DSB 2017 містить інформацію тільки про наявність раку легенів у пацієнта. Тому, в роботі [72] автори анотували знімки 754 пацієнтів тренувальної вибірки самостійно. В даній дисертації при побудові власних моделей було перевикористані ці анотовані екземпляри.

Загалом, анотовані утворення мають дуже різні характеристики в наборах даних DSB 2017 та LUNA. LUNA містить багато регіонів, де потенційно злоякісні утворення мають розмір менше за 6 мм [116], тоді як DSB 2017

містить більше утворень великого діаметру. На рис. 4.2 можна побачити розподіл діаметрів злоякісних утворень у двох наборах даних.

Як можна побачити на рисунку 4.2, в LUNA злоякісні утворення в основному знаходяться в діапазоні 5-20 мм, тоді як в DSB 2017 наборі даних міститься в основному утворення розміром від 10-40 мм.. Більше того, якщо поглянути на графік можна помітити, що на деяких зображеннях пацієнтів є злоякісні утворення розміром більше 40 мм, тоді як в LUNA таких даних немає. Щоб трохи урівняти розподіл значень в двох наборах даних, всі утворення з розміром діаметру менш як 6 мм були видалені, оскільки згідно [116] такі утворення мають малу ймовірність бути злоякісними пухлинами. Це також підтверджується, якщо перевірити розподіл підтверджених злоякісних пухлин до здорових пацієнтів, зображених на рис. 4.3.

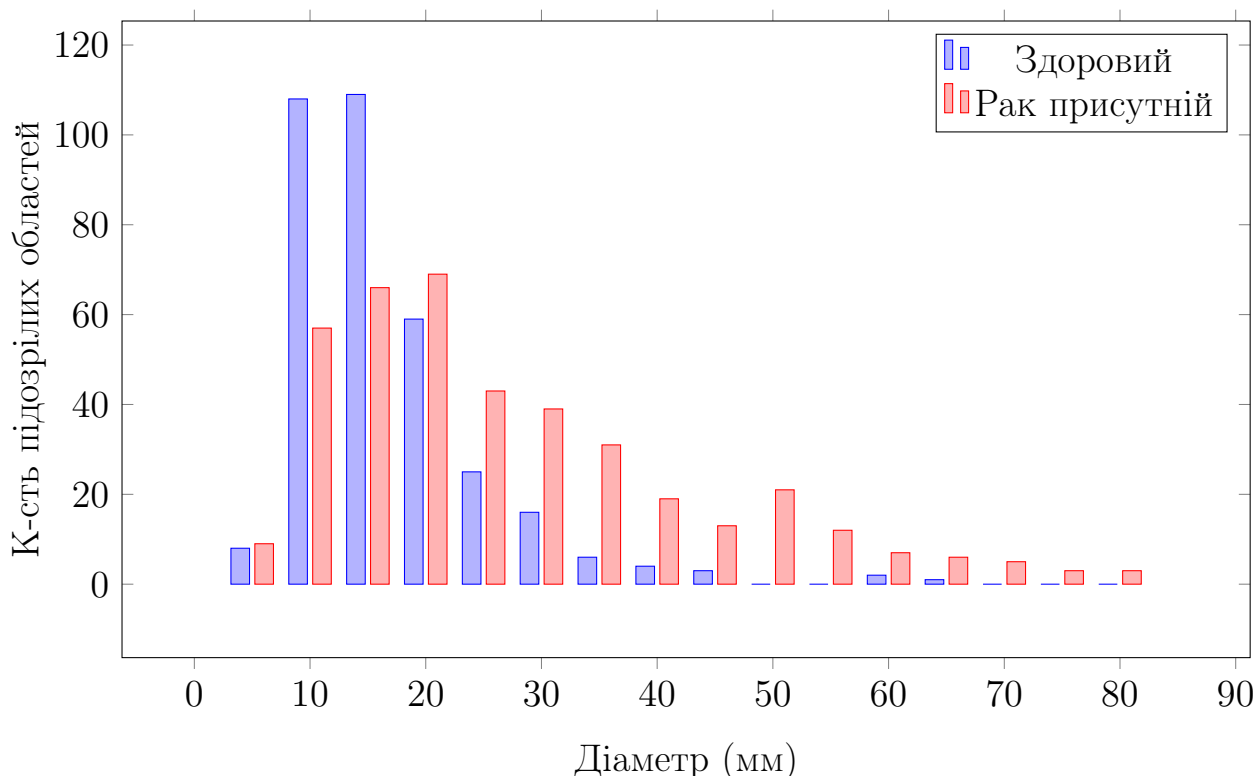


Рис. 4.3. Присутність раку в залежності від діаметру утворення на КТ знімку.

4.2. Попередня обробка КТ зображень

Значення пікселів в КТ знімках зазвичай знаходяться в шкалі одиниць Гаунсфільда. Шкала Гаунсфільда являє собою кількісну шкалу послаблення проходження рентгенівського променя через певний матеріал. Більш формально, для речовини X з лінійним коефіцієнтом послаблення μ_X денситометричні показники (англ. Hounsfield Unit) визначатимуться наступною формулою [117]:

$$HU = 1000 \times \frac{\mu_X - \mu_{\text{вода}}}{\mu_{\text{вода}} - \mu_{\text{повітря}}}, \quad (4.1)$$

де $\mu_{\text{вода}}$ та $\mu_{\text{повітря}}$ визначають лінійні коефіцієнти послаблення рентгенівського випромінювання дистильованої води та повітря в стандартних умовах. Формула сформована таким чином, що HU дистильованої води буде дорівнювати нулю. Середні денситометричні показники для різних видів речовин приведені в таблиці 4.1.

Всі виробники сканерів рентгенівської комп'ютерної томографії налаштовують та калібрують свої сканери точно вимірювати значення HU . По замовчуванню, зображення в наборі даних DSB 2017 знаходяться не в діапазоні шкали одиниць Гаунсфільда, а в у значеннях відносної рентгенопрозорості (англ. relative radiodensity). Тому перед початком роботи з КТ зображеннями необхідно перевести їх в необхідний діапазон інтенсивностей. Це можна зробити використавши значення інтенсивностей пікселів зображення, параметру “rescale slope” та значення параметру “rescale intercept” [119], що знаходяться в заголовку файлу dicom [119]:

$$HU = pixel_value \times rescale_slope + rescale_intercept, \quad (4.2)$$

де $pixel_value$ - значення пікселя в повному зображенні, отриманим сканером; $rescale_slope$ та $rescale_intercept$ - параметри масштабування отриманих значень інтенсивності рентгенівський променів. Ці значення зада-

Таблиця 4.1

**Середні денситометричні показники різних речовин в шкалі
одиниць Гаунсфілда [118]**

Речовина	Одиниця Гаунсфілда
Повітря	-1000
Легені	-500
Жир	від -100 до 50
Вода	0
Спинномозкова рідина	15
Нирки	30
Кров	від +30 до +45
М'язи	від +10 до +40
Сіра речовина	від +37 до +45
Біла речовина	від +20 до +30
Печінка	від +40 до +60
М'які тканини, контраст	від +100 до +300
Кістка	від +700 (губчаста речовина) до +3000 (кісткова речовина)

ються кожним виробником сканера окремо і можуть відрізнятися в залежності від виробника.

КТ зображення в LUNA перетворювати описаним вище чином не потрібно, через те що всі зображенням вже переведені в шкалу Гаунсфілда.

Якщо проглянути таблицю 4.1 можна відмітити, що різні середні денситометричні показники різних речовин в шкалі одиниць Гаунсфілда мають містити різні значення *HU* та відповідають певним тканинам людського організму на КТ зображенні. Цю особливість можна використати для очищення зображень від тканин, які не відносяться до легенів. Наприклад, на рисунку 4.4 показаний розподіл значень на одному із знімків набору даних

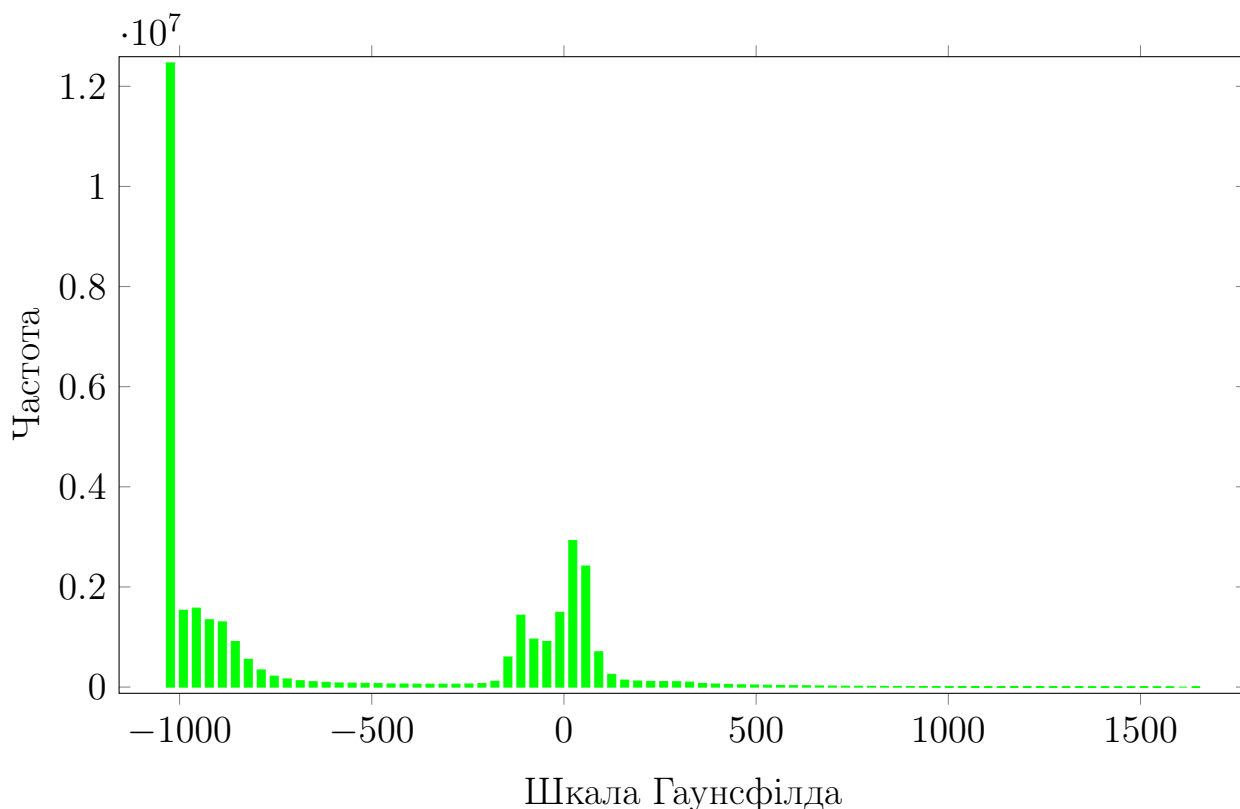


Рис. 4.4. Розподіл значень одиниць Гаунсфілда на КТ знімку пацієнта, взятого з DSB 2017

DSB 2017.

Як можна чітко побачити з рисунку 4.4, зображення КТ містить в собі інформацію про повітря, яке знаходилося в середині легень пацієнту, інформацію про тканини легенів (близько 600), кров, судини та м'які тканини. Вся ця інформація не є корисною при проведенні діагностики, тому що пухлина має знаходитися в середині легенів, а інформація про інші тканини не є репрезентативною.

Для покращення роботи ймовірнісної моделі можна відфільтрувати з зображення те, що не відноситься до легенів. Для цього, відсіюються всі значення пікселів, що більше або дорівнювали -600. Після цього, намагаємося сегментувати легені в тривимірному зображенні грудної клітки пацієнту шляхом пошуку найбільшого об'єднаного регіону на зображенні використовуючи алгоритм [120]. Зазвичай такий найбільший вибраний регіон

відповідає легеням. Таким чином ми відсікаємо все, що знаходиться поза межами легенів та залишаємо тільки те, що відноситься до легенів.

Останнім кроком в попередній обробці є зміна розміру зображення та нормалізація значень пікселів для пришвидшення навчання нейронної мережі. Новий вихідний розмір зображення буде описуватися для кожного випадку окремо, оскільки моделі, що використовувалися в цій роботі, відрізняються один від одного і накладають додаткові вимоги і обмеження зі сторони використання пам'яті. Нормалізація значень пікселів зображення була виконана використовуючи наступну формулу:

$$new_pixel = \frac{pixel - image_min_bound}{image_max_bound - image_min_bound}, \quad (4.3)$$

де *pixel* - це поточне значення пікселя в зображенні; *image_max_bound* дорівнює приблизно 400, оскільки тканини з середнім денситометричним показником більше ніж 400 являють собою кістки людини з різними денситометричними показниками; *image_min_bound* використовується зі значеннями рівними -1000 , тому що це є середнім денситометричним показником повітря, а все що менше за це значення нас не цікавить.

4.3. Збільшення даних

Одним із способів боротьби з проблемою перенавчання нейронних мереж є підхід збільшення даних (англ. data augmentation). Це досягається шляхом використання певних заданих трансформацій на існуючих зображеннях та використання їх для тренування. Зазвичай такі трансформації не збільшують розмір тренувального набору даних, але збільшують його варіативність шляхом застосування випадково вибраної трансформації до поточного екземпляру. Трансформації задаються попередньо у довільній кількості, однак під час вибору конкретного прикладу випадковим чином буде робитися вибір між оригінальним зображенням та однією із його

трансформацій. Таким чином ми збільшуємо варіативність тренувальних даних.

В даній роботі використовуються тривимірні зображення легенів пацієнта. Тому, підхід збільшення даних є ще більш ефективним, чим якби він використовувався для двовимірних даних. Причина цьому полягає в тому, що до тривимірних даних можна застосувати значно більше трансформацій, чим до двовимірних даних. При навчанні нейронних мереж використовувалися наступні методи збільшення даних:

1. Зображення випадковим чином поверталось в межах однієї з осей на певний кут.
2. Зменшення розміру зображення на випадковий фактор від $0 \dots 0.15$ та доповнювалося нулями для отримання вихідного розміру.
3. Випадковим чином зображення зміщувалося на певну випадкову відстань (в межах 15%) в напрямку кожного із осей.
4. Випадковим чином зображення перевертається в кожному із трьох напрямів.

Такі прості техніки модифікації зображення допомогли збільшити варіативність тривимірних зображень легенів та покращити характеристики узагальненості моделі. Через те, що така нейронна мережа навчається протягом багатьох епох, кожна із трансформацій мала місце при навчанні більше ніж один раз.

4.4. Двовимірна згортка та навчання за набором зразків

Найбільш простою стратегією побудови повноцінної системи автоматичної діагностики раку легенів можна вважати систему, що працює з кожним шаром КТ зображення індивідуально. Потім, інформація з кожного шару може бути об'єднана з використанням певного попередньо заданого правила. В такому разі система працюватиме з двовимірними даними,

якими значно легше оперувати та вмістити в обмежену пам'ять графічних процесорів на відмінно від тривимірних. Додатковою перевагою такого підходу можна вважати можливість використання навчання переміщенням (англ. transfer learning). Завдяки навчанню переміщенням можна значно пришвидшити та спростити навчання нейронної мережі шляхом перевикористання двовимірних згорткових нейронних мереж навчених на популярному великому наборі даних ImageNet. Як показує практика, навчання переміщенням дає змогу значно краще ініціалізувати вагові коефіцієнти нейронної мережі, що дає змогу використати низькорівневі примітиви, що були навчені на іншому наборі даних. Такий набір даних має походити з одного й того ж домену, тобто нейронна мережа має також бути навчена на статичних зображеннях. Також при відсутності великої кількості даних, можна заморозити низькорівневі згорткові шари і навчити заново тільки високорівневі шари повнозв'язної нейронної мережі вже на своєму наборі даних. Серед попередньо навчених двовимірних згорткових нейронних мереж на наборі даних ImageNet доступні різнопланові архітектури, такі як ResNet [56], DenseNet [62] та Xception [63]. Їх можна повторно використати для побудови нашої моделі.

Набір даних DSB 2017 містить тільки одну анотацію для всього КТ зображення, що робить неможливим використання підходу описаного вище на пряму. Для навчання системи, де для набору зображення є тільки одна мітка, потрібно використовувати підхід навчання за набором зразків (англ. multi-instance learning).

Навчання за набором зразків замість того щоб розглядати кожне зображення індивідуально, розглядає певний пакет (англ. bag), що містить певний набір зображень та асоціює з таким пакетом даних тільки одну мітку. Фінальне рішення про належність всього пакету до певного класу виконуються використовуючи припущення навчання за набором зразків. Для бінарної класифікації припущення говорить про те, що якщо пакет містить

в собі хоча б один позитивний екземпляр, то весь пакет даних приймається за позитивний. Пакет даних вважається негативним тільки у випадку, коли всі екземпляри даних в пакеті є негативними [121].

Реалізація підходу багато-зразкового навчання з двомірною згортковою нейронною мережею в даній роботі базується на підході, описаній в [122] для побудови системи діагностики раку молочної залози. Однак, на відмінно від способу представленому в [122], кожен шар КТ зображення в навчаній моделі представляє собою один екземпляр в пакеті даних. Також, в навченій моделі фінальна функція втрат розраховується на основі зображень з кожного шару КТ знімку, що були пропущені через вибрану двовимірну згорткову нейронну мережу.

Якщо задавати проблему більш формально, то нехай ми маємо вхідне КТ зображення пацієнта S на вході нейронної мережі. КТ зображення S містить в собі k зображень. Метою моделі визначається задача визначення присутності раку легенів для всього скану КТ. Маючи це та припускаючи, що наша модель розглядає повне тривимірне зображення КТ, задача визначається як проблема бінарної класифікації. Для кожного шару КТ зображення, ймовірність наявності пухлини можна визначити наступним чином:

$$r_n = \sigma(\omega \times CNN(k_m) + b), \quad (4.4)$$

де σ являє собою сигмоїдну функцію; ω - ваговий коефіцієнт логістичної регресії; CNN - функція, що представляє етап прямого поширення згорткової нейронної мережі; k_m - зображення шару КТ скану пацієнта на позиції m ; b - зміщення.

Маючи ймовірність пухлини на кожному шарі КТ зображення, сумарну ймовірність наявності пухлини на знімку можна розрахувати наступною формулою:

$$p(y = 1|S, k) = \max\{r_1, r_2, \dots, r_n\}, \quad (4.5)$$

$$p(y = 0|S, k) = 1 - \max\{r_1, r_2, \dots, r_n\}, \quad (4.6)$$

де y представляє коректну мітку з набору даних.

В результаті, функція втрат для побудованої моделі може бути записана наступним чином:

$$L = \frac{1}{N} \sum_{n=1}^N (-\log(p(y_n|S, k)) + \mu \|r'_n\|_1), \quad (4.7)$$

де N відповідає кількості шарів в КТ зображенні; μ - коефіцієнт розрідженості, який представляє можливість вибрати баланс між важливістю зображення та припущенням як багато шарів містять пухлину; $\|r'_n\|_1$ - L_1 норма вихідного вектора зображення на кожному з шарів КТ зображення.

Як двовимірну згорткову нейронну мережу було вибрано DenseNet [62], оскільки ця нейронна мережа істотно зменшує кількість параметрів та показує одні з найкращих результатів на класичному наборі даних для комп'ютерного зору ImageNet. Для навчання використовувалася попередньо навчена на ImageNet згорткова нейронна мережа з глибиною 121. Така попередньо навчена нейронна мережа дозволила краще ініціалізувати вагові коефіцієнти згорткової нейронної мережі та прискорити навчання моделі. Двовимірне зображення в кожному шарі КТ скану було зменшено до розміру 200×200 . Кількість шарів для знімків пацієнтів не змінювалося, незважаючи на той факт, що кількість шарів в кожному КТ знімку є різною та залежить від роздільної здатності сканера.

Під час тренування, вихід з DenseNet для кожного шару знімку КТ було збережено та використовувалося для розрахунку функції втрат за рівнянням 4.7. Під час навчання використовувався фактор розрідженості рівний 0.0001. Після розрахунку функції втрат для кожного з шарів знімку, за допомогою алгоритму зворотнього розповсюдження (англ. backpropagation) були розраховані значення градієнтів. Для оптимізації функції втрат, використовувалася модифікація стохастичного градієнтного спуску Adam.

Під час навчання, коефіцієнт швидкості навчання (англ. learning rate) в ручному режимі змінювався в залежності від динаміки процесу тренування. Початкове значення коефіцієнту навчання використовувалося на рівні 0.00005. Процес тренування моделі був зупинений в той момент, коли функція втрат не зменшувалася протягом декількох епох.

Після закінчення процесу тренування модель показала рівень коректності роботи (англ. accuracy) на рівні 62% на валідаційній вибірці.

Основна причина, чому така модель показує такий низький рівень коректності роботи полягає в тому, що вона втрачає інформацію про тривимірну природу вхідних даних. Такі значення є важливою частинкою інформації, яка допомагає робити коректні висновки про наявність пухлини на знімку.

4.5. Рекурентна нейронна мережа з механізмом уваги

Детально інформація про механіку моделі розглядалася в розділі 3. В даному розділі описується конфігурація тренувального процесу.

В існуючу модель DenseNet + LSTM після кожного Densenet блоку було додано блоки оберненої згортки (див. рис. 3.5). Додані блоки оберненої згортки при навчанні використовуються як механізм надання моделі інформації про локацію злоякісного утворення. По суті блоки оберненої згортки виконують задачу сегментації. Такий механізм надає нейронній мережі інформацію, де знаходиться проблемний регіон та стимулює більше приділяти увагу саме на локації, що будуть задані маскою в тренувальній вибірці. Блок оберненої згортки складається з декількох послідовно з'єднаних шарів згортки 1×1 та оберненої згортки для відновлення розміру зображення. В даній моделі маска може мати менший розмір за вхідне зображення. Між відповідним Densenet блоком та шарами з блоку оберненої згортки існують додаткові зв'язки, зроблені в такий ж манері, як в нейронній мережі типу U-Net [54]. Такі додаткові зв'язки допомагають зменшити вплив

проблеми втрати інформації при виконанні операції оберненої згортки.

Для навчання створеної рекурентної нейронної мережі з механізмом уваги, необхідно також зробити додаткові кроки обробки тренувальної вибірки. В наборі даних LUNA [115] є попередні описані рентгенологами координати знаходження потенційно злоякісних утворень. Використовуючи цю інформацію, були створені бінарні піксельні маски, які використовуються для порівняння того, що повертає нейронна мережа після кожного з блоків оберненої згортки і включення цієї інформації в функцію втрат. Набір даних DSB 2017 не містить інформацію про злоякісні утворення в легенях, тож ми повторно використали описи підозрілих регіонів, що були створені в рамках роботи [72].

В якості двовимірної згорткової нейронної мережі було використано архітектуру DenseNet з глибиною 121, яка була попередньо навчена на наборі даних ImageNet. Використання завчасно навченої нейронної мережі дозволило отримати кращу початкову ініціалізацію вагових коефіцієнтів мережі, що пришвидшило процес навчання. Фінальна функція втрат для навчання визначалася як сума з ваговими коефіцієнтами між функціями втрат для задачі класифікації та задач сегментації (визначених блоками оберненої згортки). Навчання нейронної мережі відбувалося за допомогою стохастичного градієнтного спуску з початковим значенням коефіцієнту навчання (англ. learning rate) 0.0005. Впродовж навчання коефіцієнт навчання змінювався вручну, в залежності від динаміки навчання побудованої мережі. В результаті, навчена нейронна мережа показала коректність роботи системи на рівні 0.81.

4.6. Одноетапні тривимірні згорткові нейронні мережі

В рамках дисертаційної роботи були навчанні такі одноетапні тривимірні згорткові нейронні мережі як C3D [89] та 3D DenseNet [90]. Для вико-

ристання КТ зображення в цих нейронних мережах було сформовано 3D зображення легенів пацієнта шляхом поєднання шарів КТ один над одним. Кількість шарів для кожного з КТ зображень відрізняється, тому що для формування вибірки бралися зображення зі сканерів різної роздільної здатності та різних виробників. Тому всі отримані тривимірні зображення пацієнтів були зменшені до розміру $120 \times 120 \times 120$. Такий розмір дозволив отримати зображення однакового розміру та вмістити нейронну мережу в пам'яті однієї відеокарти. Тривимірне зображення легенів пацієнту після попередньої обробки можна побачити на зображенні 4.5.

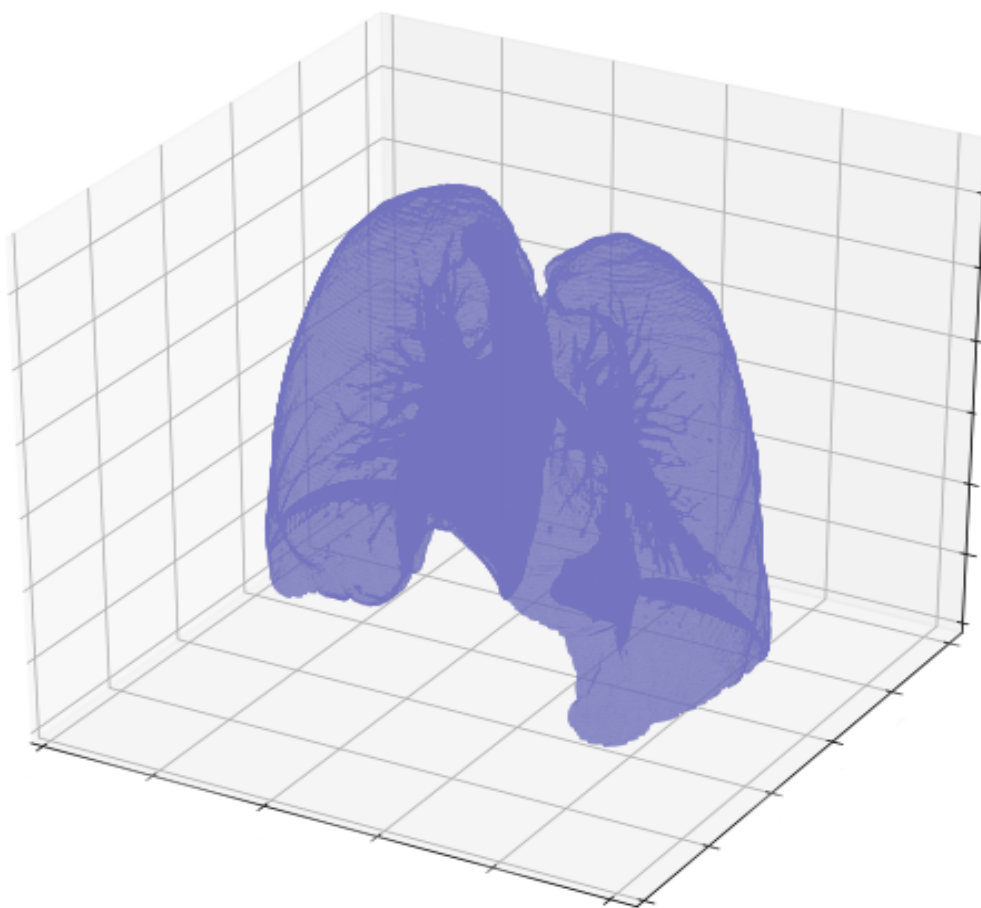


Рис. 4.5. Тривимірне зображення легенів пацієнту після сегментації та попередньої обробки. Саме в такому форматі на вхід тривимірної згорткової нейронної мережі подаються вхідні дані.

Функція втрат для тренування тривимірних нейронних мереж була за-

дана як функція перехресної ентропії таким чином, що ваговий коефіцієнт для класу представленого в меншій кількості був вищим. Навчання виконувалося з використанням модифікації стохастичного градієнтного спуску Adam, з коефіцієнтом швидкості навчання (англ. learning rate) рівним 0.0001 та 0.00004 для C3D та 3D DenseNet відповідно.

Навчена тривимірна згорткова нейронна мережа C3D показала коректність (англ. accuracy) роботи на рівні 63%. В той же час тривимірна згорткова нейронна мережа 3D DenseNet змогла показати коректність роботи на рівні 70%. Як видно з отриманих результатів, навченні нейронні мережі не показують високих показників точності прогнозування. Причиною цього є складність ознак які має вивчити нейронна мережа для більш точного прогнозування раку легенів, які можливо віднайти тільки маючи значно більшу кількість прикладів у навчальній вибірці. Для зменшення необхідної кількості даних для навчання, задачу необхідно розбити на два більш спрощених етапи - сегментація потенційно підозрілих регіонів та їх класифікація.

4.7. Підсумки отриманих результатів

Додатково до описаних експериментів було додано результати і інших публікацій, що працювали з схожими наборами даних. Це зроблено для того, щоб детально порівняти можливості кожної з архітектур та мати повну картину того, як змінюється робота системи в залежності від різних деталей доданих в пайплайн системи діагностики. Передусім ми додали до сумарної таблиці результати, що були отримані переможцями змагання Data Science Bowl в 2017 році [72]. Автори цієї роботи використовували двоетапну тривимірну згорткову нейронну мережу та отримали систему з коректністю роботи фактично рівною середньостатистичному рентгенологу. Інша робота вводить ідею системи, що складається з чотирьох етапів

[94] та можливістю враховувати інформацію з попереднього КТ зображення пацієнта. Ця робота змогла досягти трохи вищої коректності роботи за середнього рентгенолога. Всі отримані результати та результати описаних публікацій представлені в підсумковій таблиці 4.2.

Таблиця 4.2

Сумарні результати отриманні при навчанні моделей та порівняння з результатами інших робіт

Назва моделі	Набори даних	Коректність (accuracy)	AUC ROC
DenseNet + Навчання за набором зразків	Luna + DSB 2017	0.62	0.61
C3D	Luna + DSB 2017	0.65	0.66
3D DenseNet	Luna + DSB 2017	0.71	0.68
DenseNet + LSTM	Luna + DSB 2017	0.72	0.69
Рекурентна НМ описана в роз. 3	Luna + DSB 2017	0.81	0.83
Модель переможець на DSB 2017 [72]	Luna + DSB 2017	0.8	0.87
Багато-етапна згорткова нейронна мережа [94]	Luna + NLST (DSB 2017 базується на цьому наборі даних)	0.9	0.95

Згідно проведених підсумків, краще всього себе показали тривимірні згорткові нейронні мережі та рекурентна згорткова нейронна мережа з механізмом уваги. Якщо порівняти отримані результати з іншими роботами,

що також вказані у підсумковій таблиці, то можна вивести доволі чіткі закономірності в структурі таких систем. На даний момент найбільш кращим варіантом є побудова системи з використанням багатоетапних тривимірних згорткових нейронних мереж, особливо коли доступна достатня обчислювальна потужність.

Багатоетапні тривимірні згорткові нейронні мережі поділяють задачу класифікації раку легенів на декілька підзадач: сегментація вхідного медичного зображення для генерації регіонів на яких потенційно можуть знаходитися пухлини та етап класифікації, який повертає ймовірність знаходження злоякісного утворення на сегментованому регіоні. Також, в більш просунутих системах комп'ютерної діагностики додається задача сегментації легень перед сегментацією підозрілих регіонів. Таке розділення задачі допомагає спростити патерни, що має вивчити нейронна мережа для діагностики раку. Так, одна мережа вивчить патерни для виділення конкретного регіону, а інша окрема нейронна мережа знаходить ознаки, що допоможуть аналізувати отримані регіони. Таким чином замість одного складного патерну вивчаються декілька більш простих.

Запропонована комбінована архітектура згорткової рекурентної нейронної мережі змогла досягти доволі високих результатів у вигляді метрики AUC ROC на рівні 0.83. На скільки відомо здобувачеві, це найкращий результат який отримували за допомогою рекурентних нейронних мереж для задачі діагностування раку легенів. Такий результат є нижчим за отримані результати тривимірними згортковими нейронними мережами в роботах [72, 94]. Однак, перевагою побудованої рекурентної нейронної мережі є вища швидкодія, оскільки вхідні КТ знімки не проходять попередню обробку, як це робиться в системах з тривимірними згортковими нейронними мережами. Також, вища швидкодія досягається шляхом використання операції двовимірної згортки замість операції тривимірної згортки, вимоги якої до обчислювальної потужності та пам'яті ростуть квадратично з розміром

вхідних даних, а не кубічно. Також, запропонована архітектура нейронної мережі складається всього з одного етапу та однієї мережі на відмінну від систем тривимірної згорткової нейронної мережі.

Цікаво, що запропонований механізм м'якої уваги, який надає нейронній мережі інформацію про локацію злоякісних утворень під час навчання, суттєво підвищує ефективність навчання нейронної мережі. Так, без використання механізму уваги, експериментально вдалося досягнути коректності (англ. accuracy) роботи мережі на рівні 72%, тоді як при використанні механізму уваги вдалося отримати рівень коректності роботи моделі на рівні 81%. Це дуже схоже на тенденцією при використанні тривимірних згорткових нейронних мереж, де додавання підзадачі сегментації знімку легенів пацієнта значно покращує точність роботи системи.

Основні отримані результати, що представлені в даному розділі, були опубліковані в наукових роботах [5, 6, 7].

ВИСНОВКИ

Дисертаційне дослідження побудови системи комп'ютерної діагностики раку легенів показало, що створення таких систем все ще є складною задачею, рішення якої вимагає наявності достатньо великого набору прикладів зображень пацієнтів з негативним та позитивним діагнозом. Розробка медичних систем автоматичної діагностики ведеться дуже давно і до розповсюдження методів глибинного навчання найбільш популярними були багато-атласні методи. Ці методи виконували сегментацію вхідного медичного зображення використовуючи набір відомих анотованих зображень атласів, що представляли собою бінарну маску, яка відображала належність кожного пікселю до певного класу. Однак, такі системи вимагали великі обчислювальні потужності, мали проблеми з масштабуванням та не могли вивчити складні функції для проведення точної діагностики. Це стало причиною розвитку і популяризації ймовірнісних методів, що використовували глибинне навчання за основу.

В роботі було показано, що для розробки системи автоматичної діагностики раку легенів, які використовують тривимірне зображення легенів отримане за допомогою комп'ютерної томографії, можливо використати три основні підходи:

1. Для кожного шару КТ знімка застосовується згорткова нейронна мережа. Виходи мережі для кожного з шарів об'єднуються та фінальний висновок робиться на основі правил навчання за набором зразків.
2. Застосувати тривимірні згорткові нейронні мережі, які розуміють тривимірну природу вхідних даних та можуть віднайти корисні патерни використовуючи всі три просторові осі. Часто, такі системи

розділяють задачу на декілька етапів, кожен з яких використовує тривимірну згорткову нейронну мережу оптимізовану під конкретну підзадачу.

3. Використання рекурентних згорткових нейронних мереж, що використовують згорткову нейронну мережу для представлення вхідних даних в менш мірному просторі шляхом навчання менш мірного многовиду вхідних даних. Завдяки цьому на кожному шарі КТ зображення будуть виділятися тільки найбільш важливі високорівневі ознаки. Отримані ознаки оброблюються двонапрямною рекурентною нейронною мережею з вентильним вузлом (англ. *bidirection gated recurrent neural network*), яка навчається складним функціям, що описують просторові залежності та вплив між ними. Вихід рекурентної мережі повертає ймовірність наявності пухлини на знімку.

Наукова новизна отриманих результатів дисертації полягає в запропонованому здобувачем методі побудови комбінованої структури системи комп'ютерної діагностики, що поєднує двовимірну згорткову та двонаправлену рекурентну нейронну мережу LSTM. На відмінно від інших рішень, така система враховує просторові зв'язки між різними шарами знімку комп'ютерної томографії шляхом використання двонаправленої рекурентної нейронної мережі, на вході якої використовують високорівневі ознаки сформовані за допомогою двовимірної згорткової нейронної мережі. Високорівневі ознаки будуються для кожного шару знімку пацієнта. За результатами експериментів така архітектура нейронної мережі змогла досягти значення AUC ROC на рівні 0.83, що трохи нижче у порівнянні з системами тривимірних згорткових нейронних мереж, що показують значення AUC ROC на рівні 90-95%. Однак, отримані результати є найвищими результатами для рекурентних нейронних мереж, що застосовуються для побудови систем комп'ютерної діагностики раку легенів. Також, запропонована архітектура має вищу швидкодію, що досягається шляхом використання

операції двовимірної згортки замість операції тривимірної згортки, вимоги якої до обчислювальної потужності та пам'яті ростуть квадратично з розміром вхідних даних, а не кубічно.

Для ефективного навчання комбінованої структури згорткової рекурентної нейронної мережі був запропонований механізм м'якої уваги, що надав можливість нейронній мережі отримати інформацію про локацію пухлини під час навчання. Згідно проведених експериментів, такий підхід допоміг покращити показники метрики AUC ROC більш ніж на 8%.

Практичне значення отриманих результатів полягає в розширенні та удосконаленні існуючих методів побудови систем комп'ютерної діагностики. Запропонована комбінована структура згорткової нейронної мережі та двонаправленої рекурентної мережі дозволяє отримати достатньо високу точність роботи системи та підвищує точність роботи системи у порівнянні з використанням звичайних рекурентних нейронних мереж. Також, така система має менші вимоги до обчислювальних ресурсів та пам'яті за тривимірні згорткові нейронні мережі. Проведені експерименти та аналіз існуючих методів систем комп'ютерної діагностики дозволив сформулювати необхідні вимоги та підходи, які потрібно використовувати в залежності від пріоритету швидкодії чи точності роботи системи. Запропонований механізм м'якої уваги дозволяє значно підвищити ефективність навчання комбінованих архітектур згорткових рекурентних нейронних мереж, що дозволяє значно підвищити точність роботи системи.

Достовірність отриманих результатів забезпечується коректністю постановки задачі та коректним використанням математичного апарату, що теоретично обґрунтований та апробований науковою спільнотою при розробці архітектур штучних нейронних мереж. Достовірність отриманих результатів підтверджує шляхом практичного застосування отриманої системи комп'ютерної діагностики, а також порівнянням і співставленням з аналогічними існуючими системами опублікованих в інших наукових працях.

Робота над дослідженням систем автоматичної діагностики легенів може бути продовжена в майбутньому. Так, одним із основних напрямків дослідження може стати підвищення інтерпретовності моделі. Потрібно розуміти, що відповідальність за результати діагностування несе лікар, який використовує таку системи. Тому лікарям потрібно чітко розуміти, чому система прийняла рішення про наявність раку легенів у пацієнта, які області модель позначила підозрілими. Така інформація буде надзвичайно корисною для лікарів та зможе полегшити їм завдання призначення наступних процедур та більш точних методів діагностування, таких як біопсія.

Також, ще одним напрямом продовження дослідження є розробка та аналіз методів підвищення ефективності навчання, наприклад шляхом використання існуючих методів напіваавтоматичного навчання (англ. semi-supervised learning) в рамках пайплайну системи. Досягнення прогресу в цьому напрямі дозволить покращити роботу системи та зробить можливим використання системи навіть при наявності дуже малої кількості прикладів КТ зображень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] World Health Organization. (2014) World cancer report. [Online]. Available: <http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014>
- [2] World Health Organization. (2018) Cancer. [Online]. Available: <https://bit.ly/3camM1z>
- [3] Міністерство охорони здоров'я України. Чому виникає рак легень і як захистити себе. [Online]. Available: <https://bit.ly/39neL7y>
- [4] Київський міський онкологічний центр. Рак легенів: види і симптоми захворювання. [Online]. Available: <https://bit.ly/39jaeDr>
- [5] B. Chapaliuk and Y. Zaychenko, "Deep learning approach in computer-aided detection system for lung cancer," in *2018 IEEE First International Conference on System Analysis Intelligent Computing (SAIC)*, Oct 2018, pp. 1–4.
- [6] B. V. Chapaliuk and Y. P. Zaychenko, "Використання рекурентних нейронних мереж для автоматичної діагностики раку легенів," *Системні дослідження та інформаційні технології*, no. 3, pp. 33–40, 2019.
- [7] B. Chapaliuk and Y. Zaychenko, "End-to-end deep learning strategies for computer-aided lung cancer detection systems," *SCIREA Journal of Mathematics*, vol. 4, no. 5, pp. 140–155, 2019.
- [8] Bohdan V. Chapaliuk and Yuriy P. Zaychenko, "Огляд методів сегментації медичних зображень," *Системні дослідження та інформаційні технології*, no. 1, pp. 72–81, 2018.
- [9] V. Pande, P. Walters, P. Eastman, and B. Ramsundar, *Deep Learning for the Life Sciences*. O'Reilly Media, Inc., 2019.
- [10] R. Takahashi and Y. Kajikawa, "Computer-aided diagnosis: A survey wi-

- th bibliometric analysis,” *International journal of medical informatics*, vol. 101, pp. 58–67, 2017.
- [11] B. van Ginneken, B. M. ter Haar Romeny, and M. A. Viergever, “Computer-aided diagnosis in chest radiography: a survey,” *IEEE Transactions on Medical Imaging*, vol. 20, pp. 1228–1241, 2001.
 - [12] M. Firmino, A. H. F. de Moraes, R. M. Mendonça, M. L. R. Dantas, H. R. Héris, and R. A. M. Valentim, “Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects,” *BioMedical Engineering OnLine*, vol. 13, pp. 41 – 41, 2014.
 - [13] A. Criminisi, T. Sharp, and A. Blake, “Geos: Geodesic image segmentation,” in *ECCV*, 2008.
 - [14] E. Heiberg, J. Sjögren, M. Ugander, M. Carlsson, H. Engblom, and H. Arheden, “Design and validation of segment - freely available software for cardiovascular image analysis,” in *BMC Medical Imaging*, 2010.
 - [15] K. M. Pohl, J. W. Fisher, W. E. L. Grimson, R. Kikinis, and W. M. Wells, “A bayesian model for joint segmentation and registration,” *NeuroImage*, vol. 31, pp. 228–239, 2006.
 - [16] B. T. T. Yeo, M. R. Sabuncu, R. S. Desikan, B. Fischl, and P. Golland, “Effects of registration regularization and atlas sharpness on segmentation accuracy,” *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 10 Pt 1, pp. 683–91, 2007.
 - [17] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, “Automatic anatomical brain mri segmentation combining label propagation and decision fusion,” *NeuroImage*, vol. 33, pp. 115–126, 2006.
 - [18] K. O. Babalola, B. Patenaude, P. Aljabar, J. A. Schnabel, D. N. Kennedy, W. R. Crum, S. M. Smith, T. F. Cootes, M. Jenkinson, and D. Rueckert, “An evaluation of four automatic methods of segmenting the subcortical structures in the brain,” *NeuroImage*, vol. 47, pp. 1435–1447, 2009.

- [19] R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, and J. P. W. Pluim, “Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple),” *IEEE Transactions on Medical Imaging*, vol. 29, pp. 2000–2008, 2010.
- [20] Y. Hao, T. Wang, X. Zhang, Y. Duan, C. Yu, T. Jiang, and Y. Fan, “Local label learning (lll) for subcortical structure segmentation: application to hippocampus segmentation.” *Human brain mapping*, vol. 35, no. 6, pp. 2674–2697, 2014.
- [21] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, and D. Rueckert, “Leap: Learning embeddings for atlas propagation,” *NeuroImage*, vol. 49, pp. 1316–1325, 2010.
- [22] M. M. Chakravarty, P. E. Steadman, M. C. van Eede, R. D. Calcott, V. Gu, P. Shaw, A. Raznahan, D. L. Collins, and J. P. Lerch, “Performing label-fusion-based segmentation using multiple automatically generated templates.” *Human brain mapping*, vol. 34 10, pp. 2635–2654, 2013.
- [23] P. A. Yushkevich, H. Wang, J. Pluta, S. R. Das, C. Craige, B. B. Avants, M. Weiner, and S. G. Mueller, “Nearly automatic segmentation of hippocampal subfields in in vivo focal t2-weighted mri,” *NeuroImage*, vol. 53, pp. 1208–1224, 2010.
- [24] S. Nouranian, S. S. Mahdavi, I. Spadinger, W. J. Morris, S. E. Salcudean, and P. Abolmaesumi, “A multi-atlas-based segmentation framework for prostate brachytherapy,” *IEEE Transactions on Medical Imaging*, vol. 34, pp. 950–961, 2015.
- [25] L. Wang, F. Shi, G. Li, Y. Gao, W. Lin, J. H. Gilmore, and D. Shen, “Segmentation of neonatal brain mr images using patch-driven level sets,” *NeuroImage*, vol. 84, pp. 141–158, 2014.
- [26] A. J. Asman, L. B. Chambless, R. C. Thompson, and B. A. Landman, “Out-of-atlas likelihood estimation using multi-atlas segmentation.” *Medical*

physics, vol. 40, no. 4, 2013.

- [27] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “An overlap invariant entropy measure of 3d medical image alignment,” *Pattern Recognit.*, vol. 32, pp. 71–86, 1999.
- [28] R. A. Heckemann, A. Hammers, P. Aljabar, D. Rueckert, and J. V. Hajnal, “The mirror method of assessing segmentation quality in atlas label propagation,” *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1194–1197, 2009.
- [29] B. B. Avants, P. A. Yushkevich, J. Pluta, D. Minkoff, M. Korczykowski, J. A. Detre, and J. C. Gee, “The optimal template effect in hippocampus studies of diseased populations,” *NeuroImage*, vol. 49, pp. 2457–2466, 2010.
- [30] E. Konukoglu, B. Glocker, D. Zikic, and A. Criminisi, “Neighbourhood approximation using randomized forests,” *Medical image analysis*, vol. 17, no. 7, pp. 790–804, 2013.
- [31] G. Sanroma, G. Wu, Y. Gao, and D. Shen, “Learning to rank atlases for multiple-atlas segmentation,” *IEEE Transactions on Medical Imaging*, vol. 33, pp. 1939–1953, 2014.
- [32] G. Sanroma, G. Wu, M. Kim, M. A. G. Ballester, and D. Shen, “Multiple-atlas segmentation in medical imaging,” in *Medical Image Recognition, Segmentation and Parsing*. Academic Press, 2016, p. 231–257.
- [33] A. Sotiras, C. Davatzikos, and N. Paragios, “Deformable medical image registration: A survey,” *IEEE Transactions on Medical Imaging*, vol. 32, pp. 1153–1190, 2013.
- [34] Y. Ou, H. Akbari, M. Bilello, X. Da, and C. Davatzikos, “Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights,” *IEEE Transactions on Medical Imaging*, vol. 33, pp. 2039–2065, 2014.
- [35] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, “Evaluation of atlas selection strategies for atlas-based image segmentation with application to

- confocal microscopy images of bee brains,” *NeuroImage*, vol. 21, pp. 1428–1442, 2004.
- [36] A. Klein, B. Mensh, S. S. Ghosh, J. A. Tourville, and J. Hirsch, “Mindboggle: Automated brain labeling with multiple atlases,” *BMC Medical Imaging*, vol. 5, p. 7, 2005.
- [37] X. Artaechevarria, A. Muñoz-Barrutia, and C. O. de Solórzano, “Efficient classifier generation and weighted voting for atlas-based segmentation: two small steps faster and closer to the combination oracle,” in *SPIE Medical Imaging*, 2008.
- [38] J. Wan, A. Carass, S. M. Resnick, and J. L. Prince, “Automated reliable labeling of the cortical surface,” *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 440–443, 2008.
- [39] D. Zhang, G. Wu, H. Jia, and D. Shen, “Confidence-guided sequential label fusion for multi-atlas based segmentation,” *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 14, no. 3, pp. 643–650, 2011.
- [40] F. Rousseau, P. A. Habas, and C. Studholme, “A supervised patch-based approach for human brain labeling,” *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1852–1862, 2011.
- [41] H. Wang, J. W. Suh, S. R. Das, J. Pluta, M. Altinay, and P. A. Yushkevich, “Regression-based label fusion for multi-atlas segmentation,” *CVPR 2011*, pp. 1113–1120, 2011.
- [42] S. K. Warfield, K. H. Zou, and W. M. Wells, “Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 23, pp. 903–921, 2004.
- [43] A. J. Asman and B. A. Landman, “Hierarchical performance estimation in

- the statistical label fusion framework,” *Medical image analysis*, vol. 18 7, pp. 1070–81, 2014.
- [44] H. Wang, Y. Cao, and T. F. Syeda-Mahmood, “Multi-atlas segmentation with learning-based label fusion,” in *MLMI*, 2014.
 - [45] Z. Ding, X. Han, and M. Niethammer, “Votenet: A deep learning label fusion method for multi-atlas segmentation,” in *MICCAI*, 2019.
 - [46] Zhipeng Ding, Xu Han, and Marc Niethammer, “Votenet+ : An improved deep learning label fusion method for multi-atlas segmentation,” *ArXiv*, vol. abs/1911.00582, 2019.
 - [47] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
 - [48] Y. Bengio, O. Delalleau, and N. L. Roux, “The curse of highly variable functions for local kernel machines,” in *NIPS*, 2005.
 - [49] Y. Bengio and M. Monperrus, “Non-local manifold tangent learning,” in *NIPS*, 2004.
 - [50] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
 - [51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
 - [52] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in *NIPS*, 2012.
 - [53] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional

- neural networks for volumetric medical image segmentation,” *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016.
- [54] R. O., F. P., and B. T., “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, N. N., H. J., W. W., and F. A, Eds. Springer, Cham, 2015, vol. 9351.
- [55] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.
- [56] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [57] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, “Identity mappings in deep residual networks,” *ArXiv*, vol. abs/1603.05027, 2016.
- [58] K. Xia, H. Yin, and Y. Zhang, “Deep semantic segmentation of kidney and space-occupying lesion area based on scnn and resnet models combined with sift-flow algorithm,” *Journal of Medical Systems*, vol. 43, pp. 1–12, 2018.
- [59] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.
- [60] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, 2016.
- [61] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *AAAI*, 2016.
- [62] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convoluti-

- onal networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2016.
- [63] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, 2016.
- [64] H.-C. Shin, M. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, “Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1930–1943, 2013.
- [65] Y. Guo, Y. Gao, and D. Shen, “Deformable mr prostate segmentation via deep feature learning and sparse patch matching,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1077–1089, 2016.
- [66] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. rahman Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [67] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3104–3112. [Online]. Available: <http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>
- [68] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.
- [69] V. Nair and G. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *27th International Conference on Machine Learning*, 2010.
- [70] Y. LeCun, Y. Bengio, and G. Hinton, “Gradient-based learning applied to

- document recognition,” *Nature*, vol. 521, no. 7553, p. 436–444, May 1998.
- [71] T. Ross, D. Zimmerer, A. Vemuri, F. Isensee, M. Wiesenfarth, S. Bodenstedt, F. Both, P. Kessler, M. Wagner, B. M?ller, H. Kenngott, S. Speidel, A. Kopp-Schneider, K. Maier-Hein, and L. Maier-Hein, “Exploiting the potential of unlabeled endoscopic video data with self-supervised learning,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, 2018.
- [72] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, “Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3484–3495, Nov 2019.
- [73] tensorflow. tf.nn.convolution. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/nn/convolution
- [74] pytorch. torch.nn. [Online]. Available: <https://pytorch.org/docs/stable/nn.html>
- [75] mxnet. mxnet.ndarray.convolution. [Online]. Available: https://beta.mxnet.io/api/ndarray/_autogen/mxnet.ndarray.Convolution.html
- [76] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. V. Beeke, D. F. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batrah, P. Caligiuri, A. O. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. M. Hughes, A. V. Castele, S. Gupte, M. Sallamm, M. D. Heath, M. Kuhn, E. Dharaiya, R. Burns, D. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, and B. Y. Croft, “The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database

- of lung nodules on ct scans.” *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.
- [77] S. Ji, W. Xu, M. Yang, and K. Yu, “3d convolutional neural networks for human action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 221–231, 2013.
- [78] C. Olah, A. Mordvintsev, and L. Schubert, “Feature visualization,” *Distill*, 2017, <https://distill.pub/2017/feature-visualization>.
- [79] Y. T. Zhou and R. Chellappa, “Computation of optical flow using a neural network,” *IEEE 1988 International Conference on Neural Networks*, vol. 2, pp. 71–78, 1988.
- [80] M. Lin, Q. Chen, and S. Yan, “Network in network,” *CoRR*, vol. abs/1312.4400, 2013.
- [81] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *AISTATS*, 2010.
- [82] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, 2015.
- [83] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *ArXiv*, vol. abs/1502.03167, 2015.
- [84] github. C3d-tensorflow. [Online]. Available: <https://github.com/tqvinhcs/C3D-tensorflow/tree/master/pretrained>
- [85] github. I3d models trained on kinetics. [Online]. Available: <https://github.com/deepmind/kinetics-i3d>
- [86] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” *ArXiv*, vol. abs/1207.0580, 2012.
- [87] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdi-

- nov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [88] A. Geron, *Chapter 11. Training Deep Neural Networks*. O’Reilly Media, Inc., 2019.
- [89] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3d convolutional networks,” in *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [90] T. D. Bui, J. Shin, and T. Moon, “3d densely convolutional networks for volumetric segmentation,” *ArXiv*, vol. abs/1709.03199, 2017.
- [91] W. Zhu, C. Liu, W. Fan, and X. Xie, “Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification,” *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 673–681, 2018.
- [92] S. Ren, K. He, R. B. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137–1149, 2015.
- [93] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, “Mask r-cnn,” *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.
- [94] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. P. Naidich, and S. Shetty, “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nature Medicine*, vol. 25, pp. 954–961, 2019.
- [95] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn,” in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [96] B. Chapaliuk, “Overview of the three-dimensional convolutional neural networks usage in medical computer-aided diagnosis systems,” *American Journal of Neural Networks and Applications*, vol. 6, no. 2, pp. 22–28, 2020.

- [97] C. Olah. Understanding lstm networks. [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [98] B. Hammer, “On the approximation capability of recurrent neural networks,” *Neurocomputing*, vol. 31, pp. 107–123, 1998.
- [99] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Trans. Signal Process.*, vol. 45, pp. 2673–2681, 1997.
- [100] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer-Verlag Berlin Heidelberg, 2012.
- [101] Graves, “Generating sequences with recurrent neural networks,” *ArXiv*, vol. abs/1308.0850, 2013.
- [102] J. Chen and N. S. Chaudhari, “Capturing long-term dependencies for protein secondary structure prediction,” in *ISNN*, 2004.
- [103] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 1997.
- [104] T. Cukur, S. Nishimoto, A. G. Huth, and J. L. Gallant, “Attention during natural vision warps semantic representation across the human brain,” in *Nature Neuroscience*, 2013.
- [105] K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio, “Show, attend and tell: Neural image caption generation with visual attention,” in *ICML*, 2015.
- [106] V. Mnih, N. M. O. Heess, A. Graves, and K. Kavukcuoglu, “Recurrent models of visual attention,” in *NIPS*, 2014.
- [107] J. Ba, V. Mnih, and K. Kavukcuoglu, “Multiple object recognition with visual attention,” *CoRR*, vol. abs/1412.7755, 2014.
- [108] M. Malinowski, C. Doersch, A. Santoro, and P. W. Battaglia, “Learning visual question answering by bootstrapping hard attention,” in *ECCV*, 2018.
- [109] Z. Yang, X. He, J. Gao, L. Deng, and A. J. Smola, “Stacked attention networks for image question answering,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21–29, 2015.

- [110] K. Greff, R. K. Srivastava, J. Koutník, B. Steunebrink, and J. Schmidhuber, “Lstm: A search space odyssey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, pp. 2222–2232, 2017.
- [111] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, “Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans,” *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 281–284, 2017.
- [112] Оберіг. Рак легенів. [Online]. Available: <https://oberig.ua/diseases/lechenie-rak-legkogo/>
- [113] Добробут. Як проводиться лікування, в чому полягає профілактика раку легень. [Online]. Available: <https://bit.ly/3ayR7Xv>
- [114] Kaggle. (2017) Data science bowl 2017. [Online]. Available: <https://www.kaggle.com/c/data-science-bowl-2017>
- [115] A. A. A. Setio, A. Traverso, T. de Bel, M. S. N. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. H. J. Geurts, R. van der Gugten, P.-A. Heng, B. Jansen, M. Kaste, V. Kotov, J. Lin, J. Manders, A. Sónora-Mengana, J. C. García-Naranjo, M. Prokop, M. Saletta, C. Schaefer-Prokop, E. T. Scholten, L. Scholten, M. M. Snoeren, E. L. Torres, J. Vandemeulebroucke, N. Walasek, G. C. A. Zuidhof, B. van Ginneken, and C. Jacobs, “Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge,” *Medical Image Analysis*, vol. 42, p. 1–13, 2017.
- [116] H. MacMahon, D. P. Naidich, J. M. Goo, K. S. Lee, A. N. C. Leung, J. R. Mayo, A. C. Mehta, Y. Ohno, C. A. Powell, M. Prokop, G. D. Rubin, C. Schaefer-Prokop, W. D. Travis, P. E. Y. V. Schil, and A. A. Bankier, “Guidelines for management of incidental pulmonary nodules detected on ct images: From the fleischner society 2017.” *Radiology*, vol. 284, no. 1, pp. 228–243, 2017.

- [117] T. G. Feeman, *The Mathematics of Medical Imaging*. Springer, 2015.
- [118] T. M. Buzug, *Computed Tomography*. Springer-Verlag Berlin Heidelberg, 2008.
- [119] Dicomlookup. Dicom rescale intercept. [Online]. Available: <http://dicomlookup.com/lookup.asp?sw=Tnumber&q=%280028,1052%29>
- [120] C. Fiorio and J. Gustedt, “Two linear time union-find strategies for image processing,” *Theor. Comput. Sci.*, vol. 154, pp. 165–181, 1996.
- [121] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, “Solving the multiple instance problem with axis-parallel rectangles,” *Artif. Intell.*, vol. 89, pp. 31–71, 1997.
- [122] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, “Deep multi-instance networks with sparse label assignment for whole mammogram classification,” in *Medical Image Computing and Computer Assisted Intervention ? MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science*, D. M., M.-H. L., F. A., J. P., C. D., and D. S., Eds. Springer, Cham, 2017, vol. 10435.